

Discussion Paper No. 567

**PEER EFFECTS ON SUBSTANCE USE  
AMONG AMERICAN TEENAGERS**

Daiji Kawaguchi

May 2002

The Institute of Social and Economic Research  
Osaka University  
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

# Peer effects on substance use among American teenagers <sup>1</sup>

Daiji Kawaguchi<sup>2</sup>

Institute of Social and Economic Research  
Osaka University  
6-1 Mihogaoka, Ibaraki  
Osaka 567-0047 Japan.  
Tel.: +81-6-6879-8561; fax: +81-6-6878-2766  
E-mail address: kawaguch@iser.osaka-u.ac.jp

May 2002

<sup>1</sup>Due to a confidentiality agreement with the BLS on the usage of NLSY 97 GeoCode file, the part of the dataset that is constructed from the GeoCode file cannot be released.

<sup>2</sup>I am especially grateful to David Neumark for his constructive comments and continuous encouragement. I thank Jeff Biddle, Ali Berker, Frederic Durose, Jennifer Foster, John Strauss, Jeffrey Wooldridge, and the participants in the labor economics workshop at Michigan State University for helpful comments. Scott Adams and David Wetzell gave me detailed comments as well as advice on English. Donna Maurer provided careful editorial service. All remaining errors are, of course, my own.

## Abstract

The widespread use of illicit substances by American teenagers has attracted the interest of both the public and academic researchers. Among the various factors that people believe influence youth substance use, peer effects are identified as a critical determinant; substance use is considered a highly social behavior.

Identifying peer effects, however, is not an easy task. Common teenage behaviors can be due to similar unobserved characteristics of the group members or peer effects. Moreover, it is difficult to pinpoint whether a subject is affecting the group members' behaviors or vice versa.

In an attempt to overcome these difficulties, I estimate peer effects on substance usage among American teenagers using perceived peer behavior in *National Longitudinal Survey Youth 97*. School and household fixed effect estimation are also employed to ensure the robustness of the results. The data indicate robust peer effects. Moreover, the results do not change substantially in school and household fixed effect estimations.

JEL Classification: C4, I1

Key Words: Peer Effect, Substance Usage, Youth Behavior

# 1 Introduction

Widespread use of illicit substances by American teenagers attracts both public attention and research interest. The changing percentages of substance users during the 1990s are plotted in Figure 1. Although the percentage of alcohol users dropped in the early 1990s, it still remains high. This figure shows the steady trend of cigarette users at a high level.<sup>1</sup> It is also notable that the percentage of marijuana users increased from 4.4% to 9.7% between 1990 and 1997.

These figures have sparked much public interest about the reasons why teenagers use substances and what policy makers can do to reduce this usage. Besides the price of substances, peer effects or peer pressure is identified as a critical determinant, since the use of substances is considered to be a highly social behavior.<sup>2</sup>

Reacting to this interest, economists and sociologists have tried to estimate the existence and the strength of peer effects. Identifying peer effects is not easy since an observed behavior shared by a teenager and his/her peer may result from unobserved factors that group members share instead of peer effects. In addition, identifying peer effects becomes complex when the average reference group's outcome is used as a measurement of peers' behav-

---

<sup>1</sup>On the other hand, Gruber [2000] reports that the percentage of cigarette smokers increased by one-third between 1991 and 1997. (Based on youth behavior risk survey, for 9th through 12th graders, the number has increased from 27% to 36% between 1991 and 1997.)

<sup>2</sup>See Los-Angeles-Times [1999] for interviews of youth smokers on the reasons why they smoke.

ior. Determining whether a teenager’s behavior affects his/her peers or vice versa is difficult. Manski [1993] articulated this as the “reflection problem.” In addition, both current substance users and backgrounds of group members may affect individual behaviors. Although both effects are called peer effects, each has different policy implications. Distinguishing between these two effects, however, is known to be difficult (Manski [1993]).

This paper employs a critically different strategy for identifying peer effects; I identify peer effects by using teenagers’ subjective perceptions. Manski [1993] wrote:

Given that identification based on observed behavior alone is so tenuous, experimental and subjective data will have to play an important role in future efforts to learn about social effects.

Using subjective perceptions of peer behaviors, identifying peer effects is free from the problems that arise with using an average outcome as a peer variable. According to Manski [2000], this approach has not been taken seriously since he originally suggested it in 1993.

In addition, I employ school and household fixed effect estimation to ensure the robustness of my results.

## **2 Reflection problem**

Manski [1993] articulated several issues concerning the identification of peer effects using an average group outcome as a peer variable. A short sketch

of the problem follows. Let  $y$  be an outcome of interest,  $x$  be an index of an individual's reference group and  $z$  be attributes that affect the outcome directly. The outcome is characterized by

$$y = \alpha + \beta E(y|x) + E(z|x)\gamma + z\eta + u, \quad E(u|x, z) = x\delta. \quad (1)$$

If  $\beta \neq 0$ , then an individual behavior is affected by the mean of the group outcome,  $E(y|x)$ . This is the “endogenous effect.” If  $\gamma \neq 0$ , an individual behavior is affected by the group mean of the exogenous variable (background of group members); this is the “contextual effect.” If  $\delta \neq 0$ , the model exhibits the “correlated effect.” People in the same group behave similarly because their shared group characteristics are correlated with unobservable factors, such as social institutions.

Taking conditional expectation on  $z$  and  $x$ , the model becomes

$$E(y|x, z) = \alpha + \beta E(y|x) + E(z|x)\gamma + z\eta + x\delta. \quad (2)$$

To discuss the identification of parameters, we need to solve for the conditional expectation of  $y$  in terms of  $x$  and  $z$ . Using the iterated law of expectations,

$$E(y|x) = \alpha + \beta E(y|x) + E(z|x)\gamma + E(z|x)\eta + x\delta. \quad (3)$$

Solving the expression for  $E(y|x)$  and substituting it into (2), we obtain

$$E(y|x, z) = \alpha/(1 - \beta) + E(z|x)[(\gamma + \beta\eta)/(1 - \beta)] + x\delta/(1 - \beta) + z\eta. \quad (4)$$

The composite parameters are identified if  $[1, E(z|x), x, z]$  are linearly independent. This linear independence assumption fails when one of the followings occurs.

1. The individual attributes,  $z$ , is a function of a group index,  $x$ , since  $E(z(x)|x) = z(x)$ . Consequently  $E(z|x)$  and  $z$  are linearly dependent. This case occurs in the situation when group members share the same exogenous variables.
2.  $z$  does not vary with  $x$ . Since  $E(z|x)$  is constant,  $E(z|x)$  and  $z$  are linearly dependent. This case occurs if average exogenous variables are identical across groups.
3.  $E(z|x)$  is a linear function of  $x$  since  $E(z|x) = \theta x$ . This is a theoretical possibility rather than an actual possibility since  $x$  is an index for a group.

Even if the linear independence is assured, the endogenous effect ( $\beta$ ) is not identified if the contextual effect is present (i.e.  $\gamma \neq 0$ ). As Manski [1993] stressed, distinguishing the endogenous effect from the contextual effect is important because these two effects have critically different policy implications. Consider, for example, that a lecture on the health effects of smoking is provided to one class but not to the other classes in a particular school. If the lecture effectively makes students quit smoking in the class, the effect propagates to the students in the other classes through the endogenous effect. If the endogenous effect does not exist, the effect of the lecture

is limited to the class where the lecture is given. While the endogenous effect implies this “social multiplier,” the contextual effect and the correlated effect do not imply that.

Moreover, Manski [1993] warned that, in general, sample correspondence of  $E(y|x)$ , denoted by  $E_N(y|x)$ , is not identical to  $E(y|x)$ . As a result,  $\hat{\beta}$  can be calculated in the sample, even if  $\beta$  is not identified in the population. Thus, the successful calculation of  $\hat{\beta}$  does not imply anything about the identification. After all, the identification of the mixture of the endogenous effect and the contextual effect is possible under the assumption of the linear independence of  $[1, E(z|x), x, z]$ . Distinguishing between the endogenous effect and the contextual effect is, however, principally impossible.

The difficulty of the identification arises because the mean of the outcome is used as an explanatory variable. I avoid this complication in this study since the direct subjective perceptions of peers’ behavior, instead of the average behavior of peers, is the key explanatory variable. The beauty of this approach is that, once the linear independence assumption of explanatory variables is assured, the endogenous effect and the contextual effect are *separately* identified. In the next section, the model with subjective perception is discussed.

### 3 The model with perceived peer behaviors

The model with perceived peer behaviors is specified as follows:

$$y = \alpha + \beta p + E(z|x)\gamma + z\eta + u, \tag{5}$$



where  $y$  is an outcome of interest,  $x$  is attributes characterizing an individual's reference group and  $z$  is attributes that affect the outcome directly. The variable  $p$  is subjective perception of peer behaviors. Once  $E(u|p, z, x) = 0$  and linear independence of  $[1, p, z, E(z|x)]$  are assured,<sup>3</sup> the parameters,  $\alpha, \beta, \gamma$  and  $\eta$ , are consistently estimated through OLS. Since the group average of observed behaviors is not used as an explanatory variable as in Manski [1993], no complication of identification arises.

The crucial assumption in this model is that perceived rather than actual peer behaviors determine individual behaviors. This assumption is consistent with the conformist model. Akerlof [1997] argued that individuals get utility from behaving like an “average” person in a reference group. It is natural to assume that perceived peer behaviors produce the image of the “average” person. This “conformist” behavior may be reinforced through the formation of group norms. By forming group norms that require members to engage in similar behaviors, each individual group member enjoys higher utility. Becker [1996] considered a utility function ( $U$ ) that has own norm ( $N$ ) and norm of peers ( $N_P$ ) as arguments and its cross derivative to be positive ( $\partial^2 U / (\partial N \partial N_P) > 0$ ). Thus individuals can obtain higher utility through forming their own norms similar to the group norm. It is natural again to assume that each individual perceives group norms through perceived peer behaviors. Individuals behave according to their norms. To summarize, the

---

<sup>3</sup>The assumption of linear independence here is more restrictive than the assumption needed in Manski [1993], since  $p$  is newly added to the list of variables which should be linearly independent.

assumption that perceived peer behaviors determine individual behaviors is a natural one if peer effects operate through “conformist” preferences or the enforcement of group norms.

At the same time, there are many other reasons why peer effects exist. For example, American kids play baseball instead of cricket since other kids know how to play baseball, and, consequently, it is easier to find playmates. Moreover, since playing an actual game would be much more fun than playing catch, it may be an important goal for kids to find many playmates easily. In this example, utility obtained from playing baseball depends on the number of players with whom a kid plays (up to 18 players). By the same token, teenagers may obtain higher utility from tobacco if they consume it with their friends. In these situations, actual peer behaviors, rather than perceived ones, affect an individual’s behaviors. Although this possibility cannot be ruled out, it is simply assumed that an individual behavior is influenced only by the perceived peer behaviors.<sup>4</sup> Relaxing this assumption causes complication in identifying the endogenous effect as noted earlier. Thus, the interpretation of the endogenous effect in this paper should be restricted by this assumption.

## 4 Literature review

Many papers shed light on peer effects. In this section, the papers are classified by the identification strategy used. The identification of the endogenous

---

<sup>4</sup>Of course, the actual peer behaviors may determine perceived peer behaviors; however, here I assume that once perceived peer behaviors are included in the behavioral equation, the equation does not include actual peer behavior.

effect is especially focused because of its unique policy implication of the “social multiplier.” A thorough review of the existing literature reveals the strengths and weaknesses of each identification strategy. In particular, the limitation of an identification strategy that only uses observed behavior becomes clear.<sup>5</sup> I also introduce the studies that use economic theory to help provide additional insights into the identification of the endogenous effect. It should be noticed that each study does not necessarily fall into a single category since several identification strategies may be used in any given paper.

#### 4.1 Identification through proxy variable

Several studies use variables that represent “peer quality” or “neighborhood quality.” The problem of whether the endogenous effect can be identified using these proxy variables is considered. The outcome of interest (teenage pregnancy and high school dropping out in Evans et al. [1992] and high school dropping out in Crane [1991]) is characterized by

$$y = \alpha + \beta E(y|x) + z\eta + u. \quad (6)$$

Since two of studies above implicitly assume absence of the contextual effect, I also adopt that assumption. Instead of  $E(y|x)$ , however, a proxy variable,

---

<sup>5</sup>Many of studies surveyed in this section used nonlinear models such as probit. However, we discuss identification strategies in the context of linear models, since the discussions of identification fundamentally carry over. Moreover, the identification of parameters should not depend only on the nonlinearity assumption. The sizes of peer effects in those studies are discussed in terms of marginal effects.

such as the portion of students who is eligible for a school free lunch program (Evans et al. [1992]) or a neighborhood occupational structure (Crane [1991]),

$$q = E(y|x) + v \tag{7}$$

is used in the estimation. The equation actually estimated reduces to

$$y = \alpha + \beta q + z\eta + u - \beta v \tag{8}$$

The assumptions  $E(u|x, z) = 0$  and  $E(v|x, z) = 0$  are required for the identification of the endogenous effect,  $\beta$ . Assuming the absence of the correlated effect, the first assumption is assured. On the other hand, the second assumption is questionable. Since  $q$  in these studies consists of variables defined on reference groups,  $q$  are naturally functions of  $x$ . As implied by (7),  $v$  is a function of  $x$  accordingly. Thus,  $E(v|x, z) \neq 0$ . Although using proxy of  $E(y|x)$  reveals something about broadly defined peer effects, identifying the endogenous effect is impossible using a proxy variable.

## 4.2 Identification assuming the absence of the contextual effect

Many studies of peer effects fall into this category. Case and Katz [1991] looked at the effect of neighborhood average incidence on youth behavior, including drug use. Sacerdote [2000] looked at the effect of randomly assigned roommates in a college dormitory on student outcomes such as GPA. Gaviria and Raphael [2001]) looked at peer effects within schools on students'

behavior.<sup>6</sup>

These studies try to identify the structural parameters in (2). As we have seen in (4), the endogenous effect ( $\beta$ ) cannot be identified without assuming absence of the contextual effect (i.e.  $\gamma = 0$ ). Consequently, all of these studies assume the absence of the contextual effect except for Sacerdote [2000]. Replacing the population mean  $E(y|x)$  with the sample average  $E_N(y|x)$ , the (2) is estimated assuming  $\gamma = 0$ . The assumption  $\gamma = 0$  is not severe if the research interest lies in examining broadly defined peer effects. However, if the existence of the “social multiplier” is the main concern, the assumption is restrictive, since all of the observed peer effects are attributed to the endogenous effect by assumption. Thus the identification based on  $\gamma = 0$  is not appropriate for deriving policy implications since the impact of policy intervention depends on the existence of the “social multiplier.”

Most of the studies mentioned earlier were concerned about the violation of  $E(u|x, z) = 0$  and the consequent bias in the OLS estimator. Unobserved group characteristics (e.g. teachers’ competence in a school) make  $E(u|x, z)$  a function of  $x$  (the correlated effect). This correlation of  $x$  and  $u$  makes the OLS estimator of  $\beta$  biased since  $E(y|x)$  is also a function of  $x$ . Alternatively, unobserved individual characteristics correlated with observed characteristics make  $E(u|x, z)$  be a function of  $z$ . This correlation makes the OLS estimator of  $\eta$  biased and the bias could be transmitted to the estimator of  $\beta$ . For

---

<sup>6</sup>The behaviors considered in Gaviria and Raphael [2001] are substance uses, church attendance, and dropping out of high school

the latter concern, IV estimation implied by (3) solves the problem. Under assumption  $\gamma = 0$ ,  $E(z|x)$  is excluded from the behavioral equation of interest (2) but correlated with  $E(y|x)$  given  $\eta \neq 0$ ; thus  $E(z|x)$  serves as instruments for  $E(y|x)$ . An important point here is that  $u$  is allowed to be correlated with  $z$  but not with  $x$ . If  $u$  is correlated with  $x$ ,  $E(z|x)$  is correlated with  $u$  and cannot be an IV.

Evans et al. [1992] considered the case in which  $u$  includes parents' "willingness" to invest in their children when this willingness affects where the family resides ( $x$ ) so that the parents choose peer quality ( $E(y|x)$ ) endogenously. In this situation of endogenous sorting,  $u$  and  $x$  are correlated. To address this concern, they used varieties of  $E(z|x)$  as IVs.<sup>7</sup> As criticized in subsequent studies, this strategy is dubious because the correlation of  $u$  and  $x$  (caused by endogenous sorting) implies the correlation of  $u$  and  $E(z|x)$ . Thus  $E(z|x)$  is not ideal IVs for  $E(y|x)$ . Accordingly, their conclusion that peer effects do not exist after controlling endogenous sorting is dubious as well.

Sacerdote [2000] used random assignment of roommates in a college dorm to avoid this endogenous sorting. This natural experimental situation effectively rules out endogenous sorting. Gaviria and Raphael [2001] divided the sample into two sub-samples: families that had moved in previous two years and the families that had not moved. They argued that if endogenous sorting

---

<sup>7</sup>The metropolitan area unemployment rate, median family income, poverty rate, and the percentage of adults who completed college were used as IVs.

is widespread, estimates of peer effects for families that had recently moved into a new neighborhood should be larger. Their findings were mixed.<sup>8</sup> Although this method does not capture endogenous sorting that had occurred more than two years earlier, the important issue here is whether peer effects are critically different across groups. This is a clever possible strategy when the random assignment of peers is not available.

### 4.3 Identification through dynamic structure

Manski [1993] introduced some studies that exploit the dynamic structure of peer effects transmission. The estimated equation is the dynamic version of (2):

$$E_t(y|x, z) = \alpha + \beta E_{t-1}(y|x) + E_{t-1}(z|x)\gamma + z_t\eta + x_t\delta. \quad (9)$$

Since peer effects do not operate contemporaneously, the estimation is free from the “reflection problem” and opens the possibility for estimating peer effects.

Studies of Biddle [1991] and Norton et al. [1998] fall into this category. Biddle [1991] analyzed peer effects in the demand for personalized license plates using state-level aggregate data. In his study,  $y_t$  is current demand of personalized license plate in state  $x$ ,  $E_{t-1}(y|x, z)$  is the demand in the last year,  $z_t$  includes state aggregate income and the number of cars in a state. State dummies are included as  $x_t$ . He implicitly assumed  $\gamma = 0$ . He found significant peer effects in the demand.

---

<sup>8</sup>Different peer effects were found in the two sub-samples for marijuana usage but not in drinking, smoking, church attendance or dropping out of high school.

Norton et al. [1998] analyzed peer effects in sixth to ninth graders' substance uses. In their study,  $y_t$  is sixth to ninth graders' current use of substances, and  $E_{t-1}(y|x)$  is average use of substances among students who attended the same primary school. The vector  $z_t$  included sets of demographic and regional characteristics. They implicitly assumed  $\gamma = 0$ . Since peer effects are defined at an earlier schooling level, the estimation of peer effects is the dynamic version. The result of estimation showed that  $\beta$  is in the neighbor of 1.

Although these estimators are immune from Manski [1993]'s criticism, one should realize that identification of  $\beta$  crucially depends on the exclusion of contemporaneous peer effect by assumptions. Once the assumptions break down, Manski [1993]'s criticism again applies to these two studies. It is not clear whether the contemporaneous peer effect can be ruled out in those two studies *a priori*.

#### 4.4 Identification through sibling method

Aaronson [1998] attempted to address the heterogeneity bias caused by household heterogeneity such as parents' "willingness" to invest in their children. He estimated the effect of neighborhood quality, measured by poverty rate or average dropout rate, on a teenager's dropping out of high school. His model specifies outcome as

$$y_{ijt} = \alpha + E(y|x)_{tj}\beta + z_{ijt}\eta + c_j + u_{ijt}, \quad (10)$$

$i, t$  and  $j$  index individual, time, and household, respectively.  $E(y|x)_{tj}$  is a



regional average and  $c_j$  denotes an unobserved household heterogeneity. Absence of the contextual effect was assumed. The sibling method differences out  $c_j$  but it also differences out  $E(y|x)_{tj}$  if it is time invariant. Thus, the identification of the peer effect,  $\beta$ , crucially depends on the time variance of  $E(y|x)_{tj}$  along with the time invariance of household heterogeneity,  $c_j$ . To obtain variation in  $E(y|x)_{tj}$ , he used a sibling sample from households that moved. As Aaronson [1998] noted, however, the residential change and change of family's unobservable components are likely to be correlated,<sup>9</sup> and consequently the time invariance assumption on  $c_j$  may be violated. Moreover, the change of  $x$  should affect individual behavior only through the change of  $E(y|x)$  (i.e., the assumptions  $\gamma = 0$  and  $E(u|x, z) = 0$  are critical.). Regardless of these restrictions, it is informative that he found significant peer effects even after controlling household heterogeneity, since the bias caused by the household heterogeneity was critical concerns in previous studies.

The sibling method is also employed in this study to identify the endogenous effect; however, time invariance of household heterogeneity is no longer needed by using subjective perceptions of peers' behaviors, since the siblings' perceptions may vary within a household. Accordingly,  $c_j$  can be differenced out without differencing out perceptions.

---

<sup>9</sup>Negative unobservable shock within a family may make the family move.

## 4.5 Identification through economic theory

So far, I have surveyed identification strategies that only use observed behaviors as information. On the other hand, there are several notable studies that identify peer effects using prior knowledge suggested by economic theory.

Neumark and Postlewaite [1998] suggested “relative income concern” as a factor to explain the rapid increase of the labor force participation (LFP) rate among U.S. married women. Constructing an economic model of “relative income concern,” they actually estimated peer effects on labor force participation among married American women. They used a sister-in-law’s LFP as the peers’ behavior and regressed a woman’s LFP on her sister-in-law’s LFP. They found significant peer effects on the LFP decision. Moreover, they directly estimated the prediction obtained from their theory. The theory predicts that a woman is more likely to be employed if her husband’s income is less than her sister-in-law’s husband’s income and the sister-in-law is not employed. This is because the woman’s household may “win” in the income race due to the woman’s additional earnings. On the other side of coin, a woman is less likely to be employed if the woman with the husband whose income is less than the sister-in-law’s and her sister-in-law is employed. This is because the household is unlikely to “beat” the sister-in-law’s household with the woman’s additional earnings. The estimated result of this “best response function” was consistent with theoretical prediction.

Munshi [2000] also avoids spurious findings by using predictions from economic theory. He analyzed the technological diffusion during the Green

Revolution in India. He used the technological characteristics that the High Yield Variety (HYV) of rice is more sensitive to unobservable land quality than is the high yield variety of wheat. As a consequence of this technological factor, economic theory predicts stronger peer (social) effects for the diffusion of wheat. This is because, roughly speaking, an agent cannot learn much from his/her neighbor's experience when the success of technology adoption strongly depends on unobserved heterogeneous factors. The data supported this prediction; he found stronger peer effects in the wheat HYV diffusion than in the rice HYV diffusion.

These studies are persuasive because they carefully examined theoretical predictions and did more than regress the individual outcomes on group outcomes. Accordingly, these studies are free from Manski [1993]'s criticism. This strategy, however, can only be applied to the situation in which sharp theoretical predictions are available.

## 5 Estimation

This review of the existing research clearly indicates the limitation of studies that use only observed behavior to identify the endogenous effect. As has been discussed, economic theory can provide precious information for the identification of these effects. In the context of substance usage by teenagers, however, obtaining such a theoretical prediction is difficult since the preferences of teenagers, which are necessary to derive behavioral predictions theoretically, are largely unknown *a priori*. I thus use perceived peer behaviors,

which are potentially error ridden, as additional information to identify the endogenous effect. The description of data, the econometric model, and the results of estimations follow.

## 5.1 Data

The data set used in this study is the National Longitudinal Survey Youth 97 GeoCode file. The sample construction is summarized in Table 1. I used the set (10) (N=6356) as a sample for the cross-sectional studies, and I applied quasi school fixed effect estimation and sibling fixed effect estimation to the set (11) (N=6312) and (12) (N=2458) respectively. In Table 1, the sample means of outcomes are tabulated. From the tabulation, we can confirm that the sample selection does not drastically change the property of sample in terms of outcomes.

The outcomes are constructed by using questions about substance use in last 30 days. The respondent who smokes/drinks more than or equal to one cigarette/drink is defined as a smoker/drinker; similarly, the respondent who uses marijuana more than or equal to once per month is defined as a marijuana user.

In order to construct the peer variables, respondents were asked about their subjective perception of their peers' behavior by the following questions:

“What percentage of kids (in your grade / in your grade when you were last in school)

(smoke/smoked) cigarettes?

(get/got) drunk at least once a month?

(have / ever)used marijuana, inhalants, or other drugs?”

Respondents were allowed to answer the questions with one of the following five categories:

1. almost none (less than 10%)
2. about 25%
3. about half (50%)
4. about 75%
5. almost all (more than 90%).

From these categories, I constructed perceived peers’ behaviors in which “almost none” was coded as 0, and “almost all” was coded as 1.

Descriptive statistics of individual substance uses and perceived peer substance uses are tabulated in Table 2. An interesting finding is that the respondents systematically overestimate peer behaviors and the degree of overestimation is not negligible. It is worth noting that this measurement error is not a problem if we assume the variable that affects the respondents’ behaviors is the perceived peer behaviors rather than the “objective” (for econometricians) peer behaviors.<sup>10</sup>

---

<sup>10</sup>The difference between the subjective measure and the objective measure of the peer behaviors implies that we cannot calculate the size of the social multiplier effect from the size of endogenous effect ( $\beta$ ), since we need to know how individuals formed their perceptions. Only with perceived peer behaviors, however, can we confirm the existence of the endogenous effect without assuming absence of the contextual effect (i.e.  $\gamma = 0$ ), as discussed before.

## 5.2 The model

The substance use by a teenager is specified as

$$y = \alpha + \beta p + E(z|x)\gamma + z\eta + u, \quad (11)$$

where  $y$  is a binary variable set to one if the individual is a substance user. The variable  $p$  is perceived peer behaviors. The vector  $z$  contains a set of student, family, school, and regional characteristics that may affect individual substance use. The variable  $x$  is an index of the group, which is “schools” in this model. The determination of perceived behavior is specified as

$$p = \theta_1 E(y|x) + z\theta_2 + E(z|x)\theta_3 + v. \quad (12)$$

I assume  $E(u|x, z) = 0$  and  $E(v|x, z) = 0$ . Inclusion of several measures of parental involvement in vector  $z$  such as participation in PTA meetings and many variables (85 total) that may affect a youth’s substance use, makes the assumption  $E(u|x, z) = 0$  plausible.

There are still potential sources of omitted variable bias. For example, state anti-drug campaigns to reduce teenager substance use is a possible omitted variable that may affect both a respondent’s and his or her peer’s substance use (“correlated effect,” using Manski [1993]’s terminology). In this situation, the violation of  $E(u|x, z) = 0$  is likely to occur through  $E(u|x, z) = f(x)$ . To reduce this possibility, I included many regional variables in  $z$  that characterize the county where the teenager lives (e.g., state cigarettes tax rate, beer tax rates, county-level poverty rate, county-level un-

employment rate, the demographic composition of the county, and other characteristics). It is still fair to say, however, that the assumption  $E(u|x, z) = 0$  can be violated. Thus, I will later relax this assumption and use fixed effect estimation.

To ensure that the OLS is an unbiased and consistent estimator, I also assume

$$E(u|x, z, v) = 0. \quad (13)$$

This means that the error term of the behavioral equation ( $u$ ) is not correlated with the error term in the equation determining perceptions of peer behaviors ( $v$ ). Under this assumption, applying OLS (11) renders an unbiased and consistent estimator of  $\beta$ ,  $\gamma$  and  $\eta$  because

$$E(u|z, p, x) = E(u|x, z, p(E(y|x), z, v)) = E(u|x, z, v) = 0. \quad (14)$$

The cell average of  $z$  within a school is used instead of  $E(z|x)$ .<sup>11</sup> Since  $u$  is heteroscedastic due to the binary dependent variable, the variance covariance matrix is calculated using the White formula.

Since the respondent's school identification number (ID) is not available in my data set, I create the (almost) school ID by matching the respondent's county ID, school size and student/teacher ratio.<sup>12</sup>

---

<sup>11</sup>Although this first step estimation may change the asymptotic distribution of the OLS estimators, it is known that under the null of  $H_0 : \gamma = 0$  the asymptotic distribution is not affected.

<sup>12</sup>School size is classified into five categories: 1-299, 300-499, 500-749, 750-999, and 1000+ students. Student/teacher ratio is classified into four categories: less than 14, 14-17, 18-21 and more than 22. The Bureau of Labor Statistics assigns these two variables to each respondent based on the confidential information held by BLS. Thus, when each

### 5.3 The results

I report the results of the above model in Table 3.

For cigarette smoking, the coefficient 0.221 means that a 10 percentage point increase in the subjective perception of the peer smoking probability increases the probability of smoking by 2.2 percentage points. This estimate is statistically significant. For alcohol drinking, the estimated coefficient is 0.311. For marijuana usage, the OLS estimate is 0.229.

The results clearly show the existence and statistical significance of peer effects. When a teenager's perception of the percentage of his/her peers who use a substance increases by 10 percentage points, the probability that he/she will use the substance increases from 2.2 to 3.1 percentage points. Although the difference in identification strategy prohibits me from serious comparison of the estimates, the estimated peer effect is comparable to the estimated effect of Gaviria and Raphael [2001] for smoking and alcohol drinking (0.150 for smoking, 0.106 for alcohol drinking and 0.254 for drug use). The absence of the contextual effect ( $\gamma = 0$ ) is not rejected through an  $F$ -test. In summary, the results of the estimation robustly show the existence of peer effects in the *causal* sense. Moreover, due to the usage of perceived peer behavior as

---

school in a county is different in either school size or student/teacher ratio, I can identify all of the schools. If all of the schools in a county share both the same school size and student/teacher ratio, I just identify the county. In the worst-case scenario, it is to assume  $E(z|x)$  is constant within a county. This mis-matching becomes serious if the variation of  $E(z|x)$  is huge within a county. However, the direction of bias in the estimator of  $\gamma$  caused by this measurement error is not clear *a priori*, since the measurement error is mean reverting (to the county mean) and not classical. The contextual effect that operates at the county level is, however, at least captured.



the key independent variable, the results show that peer effects work through the endogenous effect. This implies the existence of the “social multiplier.”

The causal interpretation depends on the assumption  $E(u|x, z, v) = 0$ . The multitude of variables in  $z$  (82 in total), however, makes this assumption realistic.<sup>13</sup>

## 6 The quasi school fixed effect estimation

Although the previous section made the best effort to assure the assumption  $E(u|x, z, v) = 0$  is correct, the school index  $x$  nonetheless may still contain some information which systematically predicts teenagers’ substance use that the regional or school characteristics included in  $z$  fail to capture. In other words, there might be regional and school factors that encourage the teenagers’ substance use that are not observed. For example, suppose a cigarette shop is located just in front of high school A. Moreover, suppose this “unobservable” makes a student in the high school 50 percentage points more likely to smoke. Then the assumption  $E(u|x, z, v) = 0$  is violated, because

$$E(u|x, z, v) = \begin{cases} 0.5, & \text{if } x = A, \\ 0, & \text{Otherwise.} \end{cases} \quad (15)$$

---

<sup>13</sup>If the instrumental variables that are sufficiently correlated with  $p$  but not correlated with  $u$  exist, we can use Hausman [1978]’s method to test endogeneity of  $p$ . In particular, specifying  $E(u|x, z, v) = \rho v$ ,  $H_0 : \rho = 0$  can be tested given  $E(u|x, z) = 0$ . However, finding such IV is prohibitively difficult.

This possibility is addressed through school fixed effect estimation, which can be represented in the equation

$$y_{ij} = \beta p_{ij} + z_{ij}\gamma_1 + E(z|x)_j\gamma_2 + a(x)_j + u_{ij}. \quad (16)$$

Here  $i$  is a subscript for an individual and  $j$  is a subscript for a school. The coefficient  $\gamma_2$  is not identified, since  $E(z|x)_j$  is invariant within a school. The random variable  $a(x)_j$  captures the school specific unobservable which affects the teenager's substance usage, such as a cigarette shop at the school gate.

Under the assumptions  $E(u|p, z, a, x) = 0$  and  $E(a|p, z, x) = 0$ , both the random effects estimator and fixed effects estimator are consistent. However, only  $E(u|p, z, a) = 0$  needs to be true for the consistency of the fixed effect estimator. This allows for a test of the assumption  $E(a|p, z, x) = 0$  using a Hausman test. Since the random effect estimator is not efficient due to heteroscedasticity, I use the robust form of the Hausman test introduced in Wooldridge [2001]. Although it is not rigorously justified, the single variable Hausman test roughly tests the null  $E(a|p, z, x) = 0$ . Under this assumption, the estimator of the coefficient for  $p$  is roughly consistent.<sup>14</sup> The results of the random, fixed effect estimation and the Hausman tests appear in Table 4. All of the Hausman tests do not reject the null of  $E(a|p, z, x) = 0$ . Thus I am in favor of using the random effect estimator because of its relatively efficient property. Peer effects for smoking, drinking and marijuana usage are 0.22, 0.31, and 0.23 respectively. These numbers are similar to the OLS

---

<sup>14</sup>This discussion is not exactly true as far as  $p$  is correlated with the other explanatory variables.

results. This is probably due to the fact that the vector  $z$  already contains enough information to capture the school “unobservable.”

## 7 The household fixed effect estimation using the sibling sample

To reinforce the previous results, I estimate the household fixed effect model using sibling samples. One concern in the previous research (Evans et al. [1992]) was the endogeneity of peer quality due to omitted household characteristics. Peer quality can be endogenous since parents who are willing to invest in their children send their children to a school with good peers. Parental care can also directly affect a child’s behaviors. This problem can be avoided through controlling household unobserved heterogeneity using the household fixed effect estimation.<sup>15</sup>

I estimate the following model:

$$y_{ij} = \beta p_{ij} + z_{ij}\gamma_1 + E(z|x)_{ij} + c_j + u_{ij} \quad (17)$$

where  $i$  is a subscript for an individual,  $j$  is a subscript for a household, and  $c$  is a household unobserved heterogeneity. The same econometric discussion from the previous section applies.

---

<sup>15</sup>The sibling method, however, does not necessarily solve the endogeneity. Since between household difference of outcomes and perceived peer behaviors are wiped away, the identification solely depends on within household variation of outcomes and perceptions. If each sibling in a household has unobserved characteristics that determine both substance use and perception of peer behaviors, the sibling estimator is still biased. If this within household unobservable plays a more important role than the between households unobservable, the sibling estimator can be more biased. However, this discussion is unlikely to apply in the context of substance use. See Bound and Solon [1999] and Neumark [1999] for possible biases in the sibling estimator of the return to education.

The results of random and fixed effects estimation and the Hausman tests appear in Table 5. All of the Hausman tests reject the null of  $E(c|p, z, x) = 0$ . Thus, the fixed effects estimator is preferred. These results are similar to the previous OLS results that appeared in Table 4. Peer effects for smoking, drinking and marijuana usage are 0.14, 0.27, and 0.21 respectively. The fixed effects coefficient estimates are smaller than those obtained from the random effects or OLS estimation due to the positive correlation of the household unobserved heterogeneity and the peer variable. Nevertheless, even after allowing for the correlation of the household heterogeneity and the peer variable through the fixed effect model, the estimated peer effects are practically and statistically significant.

## 8 Peer effects and the demographic groups

Thus far, the existence of peer effects is a robust result. Next, it is interesting to investigate the strength of peer effects within different demographic groups. With knowledge of the group in which the endogenous effects are strong, policy makers can effectively target the policy that is likely to discourage youth substance use within that group, since he/she can expect larger policy effects through the larger “social multiplier.”<sup>16</sup> To estimate the difference in the endogenous effects across groups, I assume the following model in which the strength of peer effects may depend on the different demographic

---

<sup>16</sup>This discussion assumes, though, that sensitivity to the policy (a part of  $\eta$ ) is the same across groups.

groups:

$$y = \alpha + \beta_1 p + p * z_1 \beta_2 + z\gamma + E(z|x)\eta + u, \quad (18)$$

where  $z_1$  is the part of  $z$  that defines demographic groups.<sup>17</sup> Since the exogeneity of  $p$  was not rejected in the school and household fixed effects estimation, I assume  $E(u|p, z, x) = 0$ . Under this assumption, OLS is an unbiased estimator.

The results of the regressions appear in Table 6. Some of the estimated coefficients on the interaction terms are statistically significant and the effects are not negligible. As for the marijuana usage, females are less likely than males to be affected by their peers. For the other two substances, a gender difference was not found. Fewer peer effects were found among black teenagers. For smoking behavior, peer effects among black teenagers are one-third of that found among white teenagers. The smaller peer effects among black teenagers are statistically different from zero for all substances. Hispanics are also less vulnerable to peer pressure. An expectation is that minority teenagers might not obtain as much utility as non-minority teenagers from imitating each other. Teenagers with both biological parents are less likely to be affected by their peers in their smoking and marijuana uses. As for drinking, the coefficient on the interaction term of the peer variable and “both biological parents” is not statistically significant. The first two results may imply that the teenager who does not have both biological parents present

---

<sup>17</sup>I also estimated the endogenous effect using a sub-sample of demographic groups. The results obtained were qualitatively the same as the results obtained in this section.

is more likely to depend on his peers to form his behavior. This result is consistent with Steinberg [1987]’s. The result was not obtained in Gaviria and Raphael [2001], probably because the peer variable only interacts with the single parent dummy, not with the race dummies. Since single parents are presumably more common among minority children and smaller peer effect among minority children cancels larger peer effect among single parent children.

## 9 Conclusion

The estimation of peer effects on substance usage through perceived peer behaviors shows the existence of significant peer effects. When the perceived peer substance use is doubled, the probability that a teenager will use substances increases by forty to sixty percentage points. Moreover, the endogenous effect is found to be more important than the contextual effect when explaining the peer effects on youth substance use. This finding implies that current peer behaviors, rather than peer backgrounds, determine individual behaviors. Thus, if some exogenous shock reduces a group’s substance use, this reduction propagates to other groups of youths through the endogenous effect. Hence, policy makers can expect a “social multiplier” effect in policies that discourage youth substance use.

In my model, the endogenous effect is identified when perceived peer behaviors are exogenous. To assure this exogeneity assumption, I used a rich set of controls consisting of parent, neighborhood and school characteristics.

Moreover, the robustness of the results is confirmed through the school and household fixed effect estimations. We also find that the strength of peer effect depends on the demographic group to which a teenager belongs. Peer effect is found to be large among white teenagers and teenagers without both biological parents.

Although this paper finds a robust peer effect, this study does not shed enough light on the mechanism of peer effect itself. Thus this study is still a reduced form study of peer effects. More direct research on the mechanism of peer effect is left for future research. Also, peer effects on the several other socioeconomic behaviors such as criminal or sexual activities are the topics for the future research. I hope the robust findings in this study stimulate further investigation in this field.

## References

- Daniel Aaronson. Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources*, 33(4): 915–946, 1998.
- George A. Akerlof. Social distance and social decisions. *Econometrica*, 65(5):1005–1027, 1997.
- Gary S. Becker. *Accounting for tastes*, chapter 11. Harvard University Press, 1996.

- Jeff Biddle. A bandwagon effect in personalized license plates? *Economic Inquiry*, 29(2):375–388, 1991.
- John Bound and Gary Solon. Double trouble: on the value of twin-based estimation of the return to schooling. *Economics of Education Review*, 18: 169–182, 1999.
- Anne C. Case and Lawrence F. Katz. The company you keep: The effects of family and neighborhood on disadvantaged youths. Technical report, NBER Working Paper Series May, 1991.
- Jonathan Crane. The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology*, 96(5):1226–1259, 1991.
- William N. Evans, Wallace E. Oates, and Robert M. Schwab. Measuring peer group effects: A study of teenager behavior. *Journal of Political Economy*, 100(5):966–991, 1992.
- Alejandro Gaviria and Steven Raphael. School based peer effects and juvenile behavior. *Review of Economics and Statistics*, 83(2):257–268, 2001.
- Jonathan Gruber. Youth smoking in the U.S.: Prices and policies. Technical report, NBER Working Paper Series 7506, 2000.
- Jerry A. Hausman. Specification tests in econometrics. *Econometrica*, 46(6): 1251–1271, 1978.



- Los-Angeles-Times. Smoking among youths continue to rise, 1999. January 8.
- Charles Manski. Economic analysis of social interactions. *Journal of Economic Perspective*, 68(3):115–36, 2000.
- Charles F. Manski. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542, 1993.
- Kaivan Munshi. Social leaning in a heterogeneous population: Technology diffusion in the Indian Green Revolution. Manuscript, 2000.
- David Neumark. Biases in twin estimates of the return to schooling. *Economics of Education Review*, 18:143–148, 1999.
- David Neumark and Andrew Postlewaite. Relative income concerns and the rise in married women’s employment. *Journal of Public Economics*, 70:157–183, 1998.
- Edward C. Norton, Richard C. Lindrooth, and Susan T. Ennett. Controlling for the endogeneity of peer substance use on adolescent alcohol and tobacco use. *Health Economics*, 7:439–453, 1998.
- Bruce Sacerdote. Peer effects with random assignment: Results for Dartmouth roommates. Technical report, NBER Working Paper Series 7469, 2000.

L. Steinberg. Single parents, step parents, and the susceptibility of adolescents to antisocial peer pressure. *Child Development*, 58:269–275, 1987.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. MIT press, Cambridge, MA, 2001.

Table 1 Sample construction

	N	Average incidence (Non weighted)		
		Smoking	Drinking	Marijuana use
(1) Whole sample	8984	0.162	0.185	0.086
(2) All outcomes are available	8940	0.161	0.185	0.085
(3) Demographic variables are available	8851	0.161	0.185	0.086
(4) Relationships with parent are available	8833	0.161	0.185	0.086
(5) All peer variables are available	8518	0.165	0.190	0.088
(6) School characteristics are available	7498	0.166	0.192	0.091
(7) Grade in school is available	7495	0.166	0.192	0.091
(8) Parent's HGC available	7491	0.166	0.192	0.091
(9) Variables from parent questionnaire are available	6615	0.168	0.193	0.092
(10) Proxy variables are available* (Basic analysis sample)	6356	0.168	0.195	0.092
(11) Within school duplication occurs** (School fixed effect analysis sample)	6312	0.167	0.196	0.092
(12) Siblings data are available*** (Sibling fixed effect analysis sample)	2458	0.170	0.192	0.093

*Note:*

\*Proxy variables for school quality (If experience threat, if something stolen in school, Feel safe in school.) and neighborhood quality (if any gang in neighborhood).

\*\*Since school identification number is not available, quasi-school id, which is made out of county dummy, school size, and student-teacher ratio are used as quasi-school id. Bureau of labor statistics assigns last two variables based on the school id number.

\*\*\*Siblings are determined by the identical household id.

Table 2 Descriptive statistics of substance use and subjective measure of peer's behavior by grade

Grade	Smoke last 30 days	Peer who smoke (Subjective)	Drunk last 30 days	Peer who get drunk (Subjective)	Use marijuana last 30 days	Peer who use illegal drug (Subjective)	Number of Observation
4	0	0	0	0	0	0	2
5	0.044 (0.043)	0.118 (0.055)	0.041 (0.033)	0.077 (0.039)	0	0.068 (0.038)	37
6	0.058 (0.013)	0.147 (0.013)	0.039 (0.011)	0.037 (0.007)	0.018 (0.007)	0.084 (0.010)	426
7	0.097 (0.009)	0.273 (0.009)	0.084 (0.009)	0.111 (0.007)	0.036 (0.006)	0.174 (0.008)	1294
8	0.153 (0.011)	0.375 (0.009)	0.146 (0.011)	0.202 (0.008)	0.067 (0.008)	0.257 (0.009)	1319
9	0.238 (0.013)	0.529 (0.008)	0.269 (0.013)	0.389 (0.009)	0.145 (0.011)	0.426 (0.010)	1416
10	0.260 (0.014)	0.553 (0.008)	0.329 (0.015)	0.481 (0.009)	0.142 (0.011)	0.476 (0.010)	1204
11	0.260 (0.014)	0.553 (0.008)	0.329 (0.015)	0.481 (0.009)	0.154 (0.016)	0.474 (0.010)	624
12	0.291 (0.021)	0.555 (0.011)	0.371 (0.022)	0.513 (0.012)	0.154 (0.016)	0.487 (0.060)	34
Total	0.190 (0.006)	0.427 (0.004)	0.214 (0.006)	0.303 (0.004)	0.099 (0.004)	0.333 (0.004)	6356

Note:

1. All statistics are calculated using sampling weight.
2. Standard errors of mean are in parenthesis.

Table 3 OLS estimates of incidence of substance use

	(1)	(2)	(3)
Dependent variable	<u>Incidence of cigarettes smoking in last 30 days</u>	<u>Incidence of alcohol drinking in last 30 days</u>	<u>Incidence of marijuana use in last 30 days</u>
Method of estimation	OLS	OLS	OLS
<u>Peer (Fraction)</u>			
Peer smoke	0.218 (0.017)		
Peer drunk		0.309 (0.020)	
Peer illegal drug			0.228 (0.015)
z (Other control variables)	Yes	Yes	Yes
E[z x]	Yes	Yes	Yes
(Contextual Effect)			
F-statistics for Contextual effect	1.04 (0.379)	1.01 (0.447)	1.16 (0.163)
R <sup>2</sup>	0.145	0.158	0.128
Sample size	6356	6356	6356

*Note:*

1. The vector z contains following variables:

z (Independent Variables): Dummies if respondent lives with mother, father, biological mother, biological father, foster mother or foster father. Female dummy, age, black dummy, other minority dummy, Hispanic dummy, school grade dummies (grade 5 – grade 12), [Middle/junior high] school and high school dummies, catholic school dummy, private school dummy, student/teacher ratio category dummies (3 categories), School size category dummies (3 categories), Census regional dummies, urban dummy and proxy variables for unobserved school characteristics (If experience threat, if something stolen in school, Feel safe in school.), parents back ground variables (parent born in U.S., parent speak a language other than English in home, parent was with both biological parents at age of 14), last year's household income category dummies (less than \$20,000, \$20,001-\$40,000, \$40,001-\$60,000, \$60,001-\$80,000, more than \$80,001, and household income not available.), household size, number of household member less than age 18 and less than 6, proxy variables for parent's involvement in education (Often or sometimes participate in PTA activity, often or sometimes volunteer in school education) dummies for mother's education and father's education, dummy if any gang in neighborhood, county level variables (Share of white population, black population, Indian population, Hispanic population, share of population under 5 years old, 5-17 years old, 18-20 years old, 21-24 years old, 25-34 years old, 35-44 years old, 45-54 years old, 55-64 years old, 65-74 years old and 75+ years old, share of male in population). State tax rates (cigarettes tax and beer tax). There are 85 variables in total, counting each dummy of categorical variable as a variable.

2. White (heteroscedasticity-robust) standard errors are in parenthesis.

3. All of the test statistics are robust against heteroscedasticity.

Table 4 Incidence of substance use using within school duplication data.  
(Pseudo school random & fixed effect estimation.)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	<u>Incidence of cigarettes smoking in last 30 days</u>		<u>Incidence of alcohol drinking in last 30 days</u>		<u>Incidence of marijuana use in last 30 days</u>	
Method of estimation	Random-Effect	Fixed-Effect	Random-Effect	Fixed-Effect	Random-Effect	Fixed-Effect
<u>Peer (Fraction)</u>						
Peer smoke	0.218 (0.018)	0.218 (0.018)				
Peer drunk			0.310 (0.019)	0.305 (0.021)		
Peer drug					0.228 (0.016)	0.221 (0.016)
z (Other control variables)	Yes	Yes	Yes	Yes	Yes	Yes
Hausman Test F (35, 1201)		1.42 (0.014)		1.20 (0.124)		1.68 (0.000)
Single variable Hausman test t-statistics		0.680		0.740		-0.625
Sample size	6312	6312	6312	6312	6312	6312

*Note:*

1. School id is not available in data, so the (almost) school id is created out of county id, school size, and student-teacher ratio. Surveyor assigns last two variables based on confidential school id number.
2. Same control variables as in Table 4 were included. Some of the variables that do not vary within school dropped in fixed effect.
3. Heteroscedasticity robust standard errors are in parenthesis for estimated coefficient. For test statistics, p-values are in parenthesis. All the tests are robust against heteroscedasticity.

Table 5 Incidence of substance use using sibling data.  
(Household random & fixed effect estimation.)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	<u>Incidence of cigarettes smoking in last 30 days</u>		<u>Incidence of alcohol drinking in last 30 days</u>		<u>Incidence of marijuana use in last 30 days</u>	
Method of estimation	Random-Effect	Fixed-Effect	Random-Effect	Fixed-Effect	Random-Effect	Fixed-Effect
<u>Peer (Fraction)</u>						
Peer smoke	0.207 (0.029)	0.122 (0.040)				
Peer drunk			0.326 (0.033)	0.247 (0.044)		
Peer illegal drug					0.245 (0.025)	0.196 (0.034)
z (Other control variables)	Yes	Yes	Yes	Yes	Yes	Yes
E[z x]	Yes	Yes	Yes	Yes	Yes	Yes
(Contextual Effect)						
Hausman Test		1.74		1.69		1.21
F (35, 1201)		(0.000)		(0.000)		(0.096)
Single variable Hausman test		-3.830		-2.899		-2.944
t-statistics						
Sample size	2458	2458	2458	2458	2458	2458

Note:

1. The same control variables as in Table 4 were included. Some of the variables that do not vary within household dropped in fixed effect.
2. Heteroscedasticity robust standard errors are in parenthesis for estimated coefficient. For test statistics, p-values are in parenthesis. All the tests are robust against heteroscedasticity.

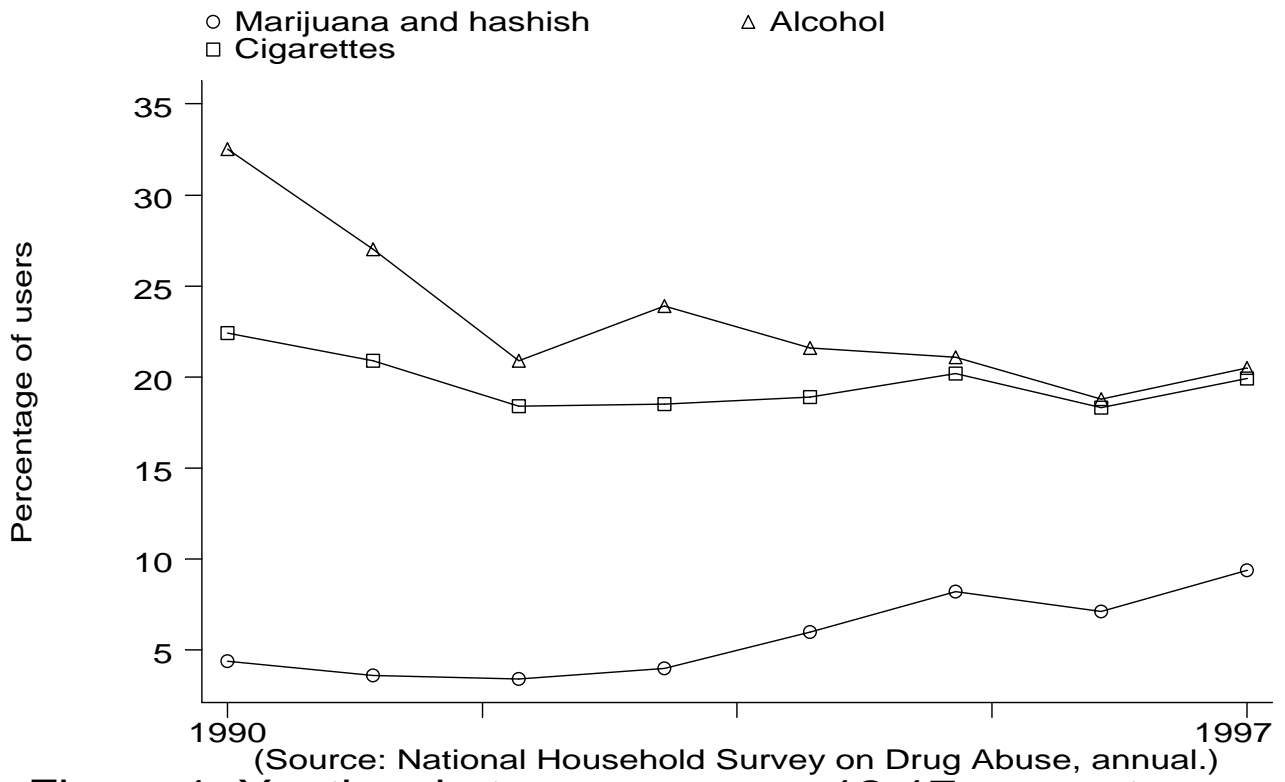
Table 6 OLS estimates of incidence of substance use

Dependent variable	(1) <u>Incidence of</u> <u>cigarettes smoking in</u> <u>last 30 days</u> OLS	(2) <u>Incidence of alcohol</u> <u>drinking in last 30</u> <u>days</u> OLS	(3) <u>Incidence of</u> <u>marijuana use in last</u> <u>30 days</u> OLS
<u>Peer</u>			
Peer's substance usage (portion)	0.417 (0.034)	0.440 (0.038)	0.352 (0.030)
Female × peer's usage	-0.014 (0.028)	-0.050 (0.033)	-0.059 (0.026)
Black × peer's usage	-0.312 (0.032)	-0.236 (0.038)	-0.175 (0.029)
Hispanic × peer's usage	-0.128 (0.039)	-0.060 (0.044)	-0.071 (0.034)
Both biological parents × peer's usage	-0.122 (0.029)	-0.027 (0.033)	-0.047 (0.027)
z (Other control variables)	Yes	Yes	Yes
R <sup>2</sup>	0.143	0.150	0.121
Sample size	6356	6356	6356

*Note:*

The same note as Table 4 applies.





**Figure 1: Youth substance use, age 12-17, current users**