HEALTH ECONOMICS
Health Econ. (2009)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hec.1488

SOCIAL INTERACTIONS AND SMOKING: EVIDENCE USING MULTIPLE STUDENT COHORTS, INSTRUMENTAL VARIABLES, AND SCHOOL FIXED EFFECTS

JASON M. FLETCHER*
Yale University, New Haven, CT, USA

SUMMARY

In this paper, I use a social interactions framework to detect whether individual smoking decisions are influenced by classmate smoking decisions. There are several large challenges in addressing this question, including the endogeneity of school (and thus classmates) through residential location choices, 'third factors' such as school-level unobservables that influence individual and classmate choices simultaneously, and the difficulty of the identification of parameters in empirical models of social interactions. In order to address these issues, I use an instrumental variables/fixed effects methodology that compares students in different grades within the same high school who face a different set of classmates and classmates' decisions. Preferred specifications suggest that increasing the proportion of classmates who smoke by 10% will increase the likelihood an individual smokes by approximately 3 percentage points. I compare these results with previous findings that are unable to use school fixed effects and/or use potentially invalid instruments and find that the current results suggest smaller social interactions in adolescent smoking decisions than some previous work. Copyright © 2009 John Wiley & Sons, Ltd.

Received 12 June 2008; Revised 9 December 2008; Accepted 2 March 2009

KEY WORDS: peer effects; social interactions; smoking

1. INTRODUCTION

Because of the severe long-term health consequences of smoking, there has been much interest in economics, health sciences, sociology, and other disciplines in documenting the determinants of adolescent smoking and suggesting policy interventions to reduce initiation and increase cessation. One primary focus of reducing adolescent smoking in the economics literature has been on increasing cigarette prices (e.g. Chaloupka and Wechsler, 1997). However, the importance of cigarette prices in youth smoking decisions has been questioned in recent work that suggests that social influences play a larger role than cigarette prices (e.g. DeCicca *et al*., 2006).

The importance of social influences on adolescent choices more generally has also been the subject of a large and growing literature in economics. Two of the areas that have received the most attention have been education outcomes and health decisions. Peer decisions and/or peer characteristics have been shown to be important in predicting elementary school achievement (Hanushek *et al*., 2003; Hoxby, 2000; Lavy and Schlosser, 2007), middle school achievement (McEwan, 2003; Summers and Wolfe, 1977; Lavy and Schlosser, 2007), enrollment in college (Fletcher, 2006), and achievement during college (Sacerdote, 2001; Zimmerman, 2003; Fletcher and Tienda, 2008). Regarding health decisions, there is

*Correspondence to: Yale University, 60 College Street, #303 New Haven, CT 06510, USA. E-mail: jason.fletcher@yale.edu

evidence that peers influence individual decisions to use drugs (Gaviria and Raphael, 2001; Kawaguchi 2004), drink alcohol (Kremer and Levy, 2003; Fletcher, 2008; Kawaguchi, 2004; Lundborg, 2006), and initiate sex (Fletcher, 2007). Several authors have also found evidence of peer influences in adolescent smoking decisions (Powell *et al*., 2005; Gaviria and Raphael, 2001; Clark and Loheac, 2007; Lundborg, 2006).

Unfortunately, credibly estimating peer influences on individual health decisions is fraught with difficulty, and most previous work has been unable to address critical challenges in identifying a primary parameter of interest – *endogenous social effects*, where an individual's propensity to behave in some way varies with the behavior of the group (Manski, 1993). Determining this effect is policy relevant because it suggests that (1) the composition of peer groups are important determinants of adolescent smoking decisions and (2) interventions that decrease the smoking propensities of individuals will decrease the smoking propensities of their peers.

There are several large challenges in addressing whether endogenous effects exist for a particular behavior. First, peer groups (assumed to be classmates in this paper) are endogenously determined by parental choices (e.g. residential choices, private school enrollment).[1] Second, there are likely to be group-level unobservables (called *correlated effects* by Manski, 1993) that influence individual and classmate choices simultaneously. Third, linear empirical models of social interactions suffer from identification problems for the coefficient that reflects endogenous effects (Manski, 1993; Brock and Durlauf, 2001), and instrumental variables or other assumptions are needed for identification.

In order to address these issues, I use the National Longitudinal Study of Adolescent Health (Add Health) and an instrumental variables/fixed effects methodology that compares students in different grades *within the same high school* who face a different set of classmates and classmates' decisions. I use school-level fixed effects to eliminate the majority of group unobservables as well as reduce the problem of the endogeneity of school (McEwan, 2003; Hoxby, 2000; Hanushek *et al*., 2003).[2] I then use several alternative sets of instrumental variables that have been suggested in previous research in order to identify the endogenous effect (Powell *et al*., 2005; Gaviria and Raphael, 2001; Fletcher, 2006). I provide evidence that previous research that was unable to use the combined IV/FE approach likely overestimate the endogenous effect of smoking decisions for adolescents. Additionally, I find evidence that for the dataset used in this paper, many instruments used in previous research on social interactions in smoking are likely inappropriate (e.g. schoolmates' family income levels, racial composition of the school).

2. BACKGROUND

There are several related research strains in the economics literature that examines the determinants of smoking choices. There is a large literature examining the importance of cigarette prices on adolescent demand. While most studies estimate that the price elasticity of demand is between -0.5 and -2.0, several studies have found no relationship between cigarette prices and adolescent demand (see Powell *et al*., 2005 and included citations). Additionally, recent work in the area has presented evidence that social norms of smoking (or 'anti-smoking sentiment') may be a more powerful predictor of smoking choices than prices (DeCicca *et al*., 2006). In particular DeCicca *et al*. find that once state-level anti-smoking sentiment is controlled, state-level cigarette prices are no longer a statistically or economically

[1]The assumption of the relevant reference group is also an important difficulty with social interactions research. I follow the literature in assuming that classmates are a relevant reference group, although other researchers have assumed larger types of reference groups (e.g. city blocks by Case and Katz, 1991).

[2]If all families choose schools based on time-invariant school characteristics, then controlling for school fixed effects controls for the main source of selection into schools.

DOI: 10.1002/hec

significant determinant of youth smoking participation.[3,4] This work suggests that examining other determinants of adolescent smoking (besides price) could have relevance for suggesting new policies to decrease adolescent smoking.

One promising direction in uncovering critical determinants of teenage smoking decisions is examining social influences. In fact, several researchers have found evidence that peer characteristics and/or peer decisions are important predictors of adolescent smoking decisions, but there are important limitations to many current studies.

Before discussing relevant research that examines the importance of social influences on smoking behavior, it is necessary to outline relevant definitions in the social interactions literature. Manski (1993, 2000) distinguishes among the following types of social effects: endogenous effects, contextual effects, and correlated effects. *Endogenous effects* occur when the propensity of an individual to behave in some way varies with the behavior of the reference group. *Contextual effects* (also called exogenous effects in the sociology literature) occur when the propensity of an individual to behave in some way varies with the exogenous characteristics of the reference group. *Correlated effects* occur when individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments. Consider the case of smoking. Endogenous effects can occur if an individual is more likely to smoke if his classmates smoke – that is, if their decisions are interdependent. Contextual effects can occur if an individual is more likely to smoke if he or she is surrounded by classmates from single-family households. Correlated effects, which are not social in nature, can occur if individuals in the same school choose to smoke because they each face low prices.

Distinguishing among these categories of effects (i.e. endogenous, contextual, and correlated effects) is important for several reasons, especially because of the implications for potential policies. Policies that take advantage of endogenous effects are likely to produce a *social multiplier*. For example, if smoking in schools is subject to endogenous effects, a policy that decreases the propensity to smoke of an individual or a group of individuals within a school will affect other individuals who were not directly targeted in the policy – the effect of the policy is multiplied through social interactions. On the other hand, contextual changes may not imply the same multiplier effect responses to an exogenous shock. For example, adding higher income students to a low-income school would benefit the students in the receiving school, but the students at the sending school would be worse off. Thus, the gains to the former school would offset the losses from the latter school so that there would be no predicted aggregate multiplier effect.[5] Finally, correlated effects are not social in nature, and failure to adequately control for these effects can lead to spurious conclusions about the importance of social influences on individual choices.

There is a large and growing literature that attempts to distinguish among the types of effects and establish the importance of social and non-social effects on smoking decisions. I only discuss the papers most related to the current paper. First, Gaviria and Raphael (2001) use the 10th grade responses from a school-based US longitudinal survey (NELS) and an instrumental variables approach to find that moving a child from a school with no smokers to a school with 25% smokers increases the individual's probability of smoking by nearly 4 percentage points. Although Gaviria and Raphael present results from the relevant statistical tests in a two-stage least squares framework (*F*-statistics, over-identification tests), it is unclear if *any* of their instruments are appropriately excluded from the main equation.

[3]The authors do find some mixed evidence on the effects of price on the conditional demand of cigarettes. The authors also point out that other researcher have questioned whether higher prices are an effective youth anti-smoking policy (Douglas, 1998; Douglas and Hariharan, 1994; DeCicca *et al.*, 2002).

[4]There is also some suggestion that cigarette smuggling across state borders may not allow additional increases in cigarette taxes to be as effective (Luccassen *et al.*, 2005).

[5]Hoxby (2000) and others have pointed out that contextual effects that are non-linear could imply distributional consequences of changing the composition of schools.

DOI: 10.1002/hec

They use the proportion of classmates with parents with drug problems, the proportion of classmates with college-educated parents, the proportion of classmates from single-parent families, and other measures.[6] While this paper was one of the first to attempt to identify endogenous effects in teenage risky behaviors, the authors were unable to control for group-level unobservables or the endogeneity of school with school fixed effects because the sample was a single cohort of US students.[7]

A paper by Powell *et al.* (2005) also uses a school-based survey and an instrumental variables methodology and finds that moving a student from a school where no children smoke to a school were 25% of the children smoke increases the probability the individual smokes by over 14 percentage points. Although Powell *et al.* use several sets of instruments and employ many statistical test to access the validity of their two-stage least squares estimates (e.g. *F*-stats, over-identification tests), their tests suggest that peer-level smoking decisions are exogenous, and it is not clear that any of their instruments are valid, which reduces their ability to rely on over-identification tests.[8]

There are two papers that are closer in spirit to the current paper. Lundborg (2006) uses a cross section of Swedish teenagers to examine smoking decisions (as well as drinking and illicit-drug use). Importantly, the author controls for school/grade fixed effects and uses variation in peer behavior across classes within schools and grades.[9] The author then uses average classmate characteristics such as proportion living in a single-parent household, proportion born outside Sweden, and proportion living in an apartment as instruments to identity the endogenous effect of smoking. While these instruments would be questionable in a cross-sectional analysis without fixed effects (since they would likely be correlated with unobserved school-level attributes), using school/grade fixed effects reduces this concern. For his Swedish data, Lundborg (2006) finds that moving a child from a classroom with no smokers to a classroom with 25% smokers increases the probability of smoking by over 12 percentage points. While the author argues that there is no sorting within grade level in Sweden, using school/grade fixed effects is less appropriate in the US context because there is evidence that individuals are non-randomly assigned to classes within grades within schools (e.g. Vigdor and Nechyba, 2004; Hoxby, 2000).

Finally, Clark and Loheac (2007) use a similar strategy as the current paper. They use the Add Health dataset and school fixed effects, but they identify the endogenous social effects through the use of lagged values of peer smoking decisions rather than contemporaneous decisions. In principle, using lagged values of the peers' choices addresses the simultaneity of group- and individual-level decisions (termed the 'reflection problem' by Manski, 1993) because current behavior cannot impact past peer decisions, but the approach is not without problems. Importantly, the researcher must know or assume the appropriate lag length (Manski, 1995, 2000).[10,11] To the extent that lagged peer choices and contemporaneous peer choices are highly correlated, and the lag length is mis-specified (for example, the 'true' time period for endogenous social effects is contemporaneous peer behaviors), the lagged peer choices could become a proxy measure of contemporaneous peer choices and without additional

[6]Gaviria and Raphael (2001) and others argue that this approach is valid because, '…students are less exposed to the family background of their school peers than they are exposed to the family background of peers residing in the same neighborhood'.

[7]The authors compared individuals who were recent movers versus those who were immobile to examine the potential bias from the endogeneity of school and found relatively large differences – a difference of nearly 50% in the estimated endogenous effect coefficient in the case of smoking.

[8]They use the racial composition of the high school, proportion of classmates with college-educated parents, Census-based income, Census-based density, proportion of classmates with married parents, and other measures. In most cases, it is easy to construct examples that undermine the validity of these instruments. For example, the racial composition of the school is likely correlated with unmeasured school resources that are in turn correlated with tobacco policies, availability, etc.

[9]McEwan (2003) uses a similar approach with data from Chile on educational achievement for middle-schoolers.

[10]Manski (1995, p. 136) states, 'Of course one cannot simply specify a dynamic model and claim that the problem of inference on social effects has been resolved. Dynamic analysis is meaningful only if one has reason to believe that the transmission of social effects follows the assumed temporal pattern'.

[11]Clark and Loheac note that they must also assume that behavior is not in a steady-state equilibrium in order to use their approach.

DOI: 10.1002/hec

restrictions (e.g. instrumental variables) may not be properly identified.[12,13] Because the authors use peer measures of behavior that are separated by gender, it is difficult to compare the magnitude of their results with previous research, but for smoking they generally find small endogenous social effects.[14]

In this paper, I build on previous research to provide further evidence on the importance of social influences on adolescent smoking decisions. Importantly, I use instruments that are more credibly excluded from the main equation predicting smoking than have been used in much previous research. Additionally, I use school-level fixed effects in order to overcome the endogeneity of school as well as unobserved school-level characteristics that could jointly determine individual and group-level smoking outcomes. Instead of using school/grade fixed effects like Lundborg (2006), I use classmates within a grade to construct the relevant peer measure. This peer measure sidesteps the problem of selection within grade, which is relevant in the US context, at the expense of using a broader measure of the relevant peer group (grade level rather than classroom level).[15]

3. DATA

The data in this study come from the restricted version of the National Longitudinal Study of Adolescent Health (Add Health). Add Health is a school-based, longitudinal study of the health-related behaviors of adolescents and their outcomes in young adulthood. Beginning with an in-school questionnaire administered to a nationally representative sample of students in grades 7 through 12 in 1994–1995, the study follows up with a series of in-home interviews of students approximately 1 year and then 6 years later. Other sources of data include questionnaires for parents, siblings, fellow students, and school administrators. By design, the Add Health survey included a sample stratified by region, urbanicity, school type, ethnic mix, and size. Pre-existing databases (e.g. census data) have been linked with the individuals in the sample and provide information about neighborhoods and communities.[16]

Of the over 19 000 individuals who completed the Wave 1 survey, 18 182 have cross-sectional weights, valid school identification codes, and reported whether they smoked cigarettes in Wave 1 of the survey. I restrict the sample to individuals in schools with 12th grade (eliminating approximately 4000 middle school students in schools that end at 8th or 9th grade) and drop those students where the number of sampled individuals in the school-grade was fewer than 20, leaving 13 602 students. Non-response to some of the family, individual, or school-level characteristics leaves an analysis sample of 13 381.[17] Unweighted summary statistics are presented in Table I.

I include individual-, family-, and school-level variables in the empirical models. For individual- and family-level variables, I include age, gender, race, grade level, an ability proxy (Peabody Picture Vocabulary Test score), marital status of parents, religious attendance, mother's education attainment,

[12]Clark and Loheac (2007) report that the correlation between the 1-year lagged peer smoking measure and contemporaneous measure is 0.54.

[13]A useful thought experiment is to consider whether the specifications in Clark and Loheac would be reasonable if the measures of lagged peer behaviors were taken the prior week or even prior day. Since smoking and peer group composition are highly persistent from day to day (or week to week), we may be skeptical of whether the lagged values are not acting as proxy values for current peer behaviors. This issue may still be problematic for time periods of 1 year. I thank an anonymous referee for suggesting this discussion.

[14]Increasing the proportion of male peer group members who smoke by 25% is predicted to increase own-smoking participation by 2.2%.

[15]This paper is similar in spirit to Hanushek *et al.* (2003) and Hoxby (2000) who use administrative data from Texas to examine peer effects in educational achievement. Important differences include the use of non-administrative data that include rich family-level information as well as the examination of smoking decisions rather than academic achievement. Using survey data rather than administrative data allows me to have broad geographic coverage, increasing the generalizeabilty of the results, as well as the ability to examine risky behaviors that are not included in school administrative data.

[16]See Udry (2003) for full description of the Add Health data set.

[17]In order to keep 3000 students, I impute family income and maternal education and create a dummy variable for missing parental data. I also impute values for whether a household member smokes for 2000 respondents.

DOI: 10.1002/hec

Table I. Summary statistics: Add Health data wave 1, $N = 13\,318$

Variable	Mean	Std.	Min	Max
Smoke	0.27	0.45	0	1
7th Grade	0.03	0.18	0	1
8th Grade	0.04	0.19	0	1
9th Grade	0.21	0.41	0	1
10th Grade	0.26	0.44	0	1
11th Grade	0.25	0.43	0	1
12th Grade	0.21	0.41	0	1
Age	17.72	1.39	13	22
Male	0.50	0.50	0	1
White	0.52	0.50	0	1
Black	0.21	0.40	0	1
Hispanic	0.18	0.38	0	1
PVT score	100.51	14.62	13	138
Household Member smoke[a]	0.44	0.50	0	1
Maternal education[a]	13.14	2.30	0	21
Family income ($1000s)[a]	0.46	0.43	0	10
Number of older siblings	0.88	1.25	0	14
Older sibling dummy	0.51	0.50	0	1
Religious attendance	26.10	21.27	0	52
Married household	0.60	0.49	0	1
Missing parent information dummy	0.33	0.47	0	1
Rural	0.24	0.43	0	1
Urban	0.35	0.48	0	1
Classmate-level characteristics				
% Classmates smoke	0.27	0.13	0	0.7
Number of sampled classmates	113	162	20	576
% Classmates male	0.50	0.07	0.28	1
% Classmates black	0.19	0.23	0	1
% Classmates Hispanic	0.18	0.22	0	0.9
Mean family income ($1000s)	45.45	16.50	15	150
Mean maternal education level	13.14	0.89	10.3	16.2
% Classmates with older sibling	0.50	0.09	0.2	0.8
Mean number of siblings	0.87	0.27	0.2	2.1
% Classes whose household member smokes	0.44	0.14	0	0.88
% Classmates married households	0.60	0.14	0.17	1
Mean religious attendance	26.05	7.43	7.7	52

[a]Imputed variable.

family income, rural status, the number of siblings, and whether a household member smokes. Peer-level characteristics include proportion male, racial distribution, average religious attendance, average family income, the number of individuals sampled in the grade, mean number of siblings, mean number of older siblings, proportion with household members who smoke, and proportion in a married household.

4. EMPIRICAL METHODS

The primary empirical specification for this paper is the linear-in-means model of social interactions (Case and Katz, 1991; Manski, 1993):

$$Y_{igs} = X_{igs}B + \bar{X}_{-igs}\delta + W_s\theta + \alpha\bar{Y}_{-igs} + \varepsilon_{igs} \quad (1)$$

where Y_{igs} is the smoking choice of individual i in grade g in school s, individual and family characteristics are contained in a vector X, classmate characteristics are measured as grade-level averages of the X vector excluding the individual, labeled $\bar{X}_{-igs}$.[18] School-level factors are measured in the vector, W_s, which is

[18]Hoxby (1999) points out that this procedure introduces measurement error in the measure.

DOI: 10.1002/hec

eliminated in school-level fixed effects specifications. Finally, $\bar{Y}_{-igs}$ is the grade-level average outcome excluding the individual (i.e. the proportion of individuals in the same grade and school who report smoking). The main coefficient of interest is the endogenous effect, α, which indicates the extent to which individuals are influenced by their peers' choices to smoke. If α is estimated to be positive, interventions that change the smoking behavior of individuals (or subsets of individuals) within a reference group would be predicted to spillover on non-treated individuals in the same reference group.[19]

As noted above, α in Equation (1) is not identified without the use of instrumental variables or other methods (Manski, 1993; Brock and Durlauf, 2001). I use instruments that have been used in previous research as well as a smaller set that is likely more reasonably excluded from the main equation. In all results ε_{igs} are allowed to be arbitrarily correlated across individuals in the same school.

5. RESULTS

In Table II, I present standard OLS regression results that examine the individual-level, family-level, and grade-level determinants of adolescent smoking decisions. Column 1 presents results for the full analysis sample. First, the grade-level dummy variables indicate a sharp rise in cigarette smoking between 8th and 9th grade (7th grade is omitted) that then only slightly increases during high school. As is well known, white students are more likely to report smoking – 18 percentage points more likely than black students. No statistically detectable differences are found between white and Hispanic students. Individuals in married households are less likely to smoke, and individuals with household members who report smoking are 13 percentage points more likely to smoke. Individuals with older siblings are more likely to smoke (even conditional on number of total siblings), as are those who report less frequent religious attendance. The racial distribution of a school is also found to influence the probability an individual smokes.[20]

Columns 2 and 3 in Table II stratify the baseline results by gender. Surprisingly, most associations are similar for males and females.

Table III stratifies the baseline results by race. The steep increase in smoking propensities across grade levels is shown to be confined to white students. The relationship between having a household member who smokes and student reports of smoking is highest for white students and low for black students. Living in a married household seems to only confer smoking benefits to white students, as do higher test scores. A higher proportion of classmates who attend church are associated with decreases in the propensity to smoke for Hispanic and black students but not white students.

In Table IV, I examine whether there is evidence of endogenous effects in smoking decisions.[21] Column 1 presents baseline results from previous tables for comparison. Most previous research has assumed that $\delta = 0$ in Equation (1), and then used the peer-level characteristics as instruments for the peer-level outcomes. I follow this specification in Columns 2 (main results) and 3 (first stage results). The results in Column 2 suggest that there is a large, economically important endogenous effect of peer-level smoking on individual smoking decisions. The coefficient can be interpreted to suggest that moving

[19]Previous research has used variation in multiple cohorts' peer characteristics within a school to examine the effects of peer characteristics (but not outcomes) on individual outcomes, where the coefficient of interest is δ rather than α (e.g. Hoxby, 2000; Hanushek *et al*., 2003; Lavy and Schlosser, 2007).

[20]The likely reason that many of the school-grade characteristics are not statistically significant is the high collinearity among the characteristics. The first factor of a factor analysis of the school-grade characteristics (which could be summarized as an 'advantage factor', loading on average education, average income, etc.) is statistically significant in unreported results. A one standard deviation increase in the factor is associated with a 3.3 percentage point increase in the probability of smoking.

[21]All results use 2SLS rather than two-stage probit specifications; the results of the two specifications are quite similar. A comparison of the endogenous effects coefficient estimates is available from the author.

DOI: 10.1002/hec

Table II. Individual-, family-, and peer-level determinants of adolescent smoking full sample and by gender

Outcome	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS
Sample	Full	Male	Female
8th Grade	0.010	0.008	0.021
	(0.038)	(0.042)	(0.050)
9th Grade	0.093***	0.092**	0.096**
	(0.033)	(0.035)	(0.048)
10th Grade	0.086**	0.083*	0.091*
	(0.038)	(0.045)	(0.053)
11th Grade	0.098**	0.086*	0.115*
	(0.042)	(0.047)	(0.060)
12th Grade	0.069	0.075	0.070
	(0.046)	(0.055)	(0.068)
Age	0.015**	0.031***	−0.000
	(0.006)	(0.009)	(0.008)
Male	0.003		
	(0.012)		
Black	−0.175***	−0.156***	−0.195***
	(0.016)	(0.022)	(0.023)
Hispanic	−0.021	0.013	−0.055***
	(0.017)	(0.021)	(0.019)
Other race	−0.018	0.022	−0.059*
	(0.032)	(0.044)	(0.032)
PVT score	−0.001***	−0.001	−0.001***
	(0.000)	(0.000)	(0.001)
Maternal education	0.000	0.002	−0.002
	(0.002)	(0.002)	(0.003)
Family income	0.004	−0.013	0.018
	(0.010)	(0.013)	(0.013)
Married household	−0.031***	−0.037***	−0.023**
	(0.009)	(0.013)	(0.011)
Household smoke	0.130***	0.124***	0.133***
	(0.014)	(0.014)	(0.018)
Religious attendance	−0.002***	−0.002***	−0.002***
	(0.000)	(0.000)	(0.000)
Number of older siblings	0.022***	0.017**	0.026***
	(0.004)	(0.007)	(0.006)
Number of siblings	−0.014***	−0.003	−0.025***
	(0.004)	(0.006)	(0.005)
Parent missing dummy	0.000	−0.006	0.006
	(0.009)	(0.012)	(0.013)
Rural status	0.006	0.020	−0.009
	(0.010)	(0.015)	(0.012)
Urban status	0.003	0.000	0.002
	(0.010)	(0.014)	(0.012)
% Black	−0.107**	−0.079	−0.130***
	(0.040)	(0.055)	(0.049)
% Hispanic	−0.292***	−0.263***	−0.322***
	(0.039)	(0.048)	(0.042)
% Other race	0.159	0.360	−0.047
	(0.161)	(0.282)	(0.178)
Mean family income	−0.001*	−0.001*	−0.001
	(0.001)	(0.001)	(0.001)
Mean maternal education	0.009	0.021	−0.002
	(0.012)	(0.017)	(0.014)
% Married households	−0.000	0.000	−0.001
	(0.001)	(0.001)	(0.001)
Mean number of siblings	−0.008	0.004	−0.019
	(0.018)	(0.023)	(0.018)
Mean religious attendance	−0.001	0.001	−0.003**
	(0.001)	(0.001)	(0.001)
Number sampled	0.006	0.007	0.005
	(0.005)	(0.006)	(0.006)

DOI: 10.1002/hec

Table II. *Continued*

Outcome	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS
Sample	Full	Male	Female
Constant	0.107	−0.514**	0.734***
	(0.167)	(0.229)	(0.197)
Observations	13 318	6599	6719
R-squared	0.10	0.08	0.13

Robust standard errors, ***1, **5, *10%.

a student from a school with 0% smoking to 25% smoking increases the individual's likelihood of smoking by 12 percentage points (0.481/4).[22] This result is quite consistent with Lundborg (2006) and Powell *et al.* (2005). The reported *F*-statistic is large (nearly 31) but the over-identification test strongly rejects the null of excludable instruments (p-value < 0.01). This result provides evidence that assuming no contextual effects and using the peer-level characteristics as instruments for peer-level outcomes is questionable. One reason that the instruments are questionable is that parents likely choose schools based on school characteristics (e.g. % black).[23] We might expect that this endogeneity bias likely increases the estimated association between the peer outcomes and individual outcomes – biasing upward the endogenous social effect, α. Additionally, there could be school-level unobservables, such as 'Do Not Smoke' campaigns that simultaneously decrease peer choices to smoke as well as individual choices to smoke. These omitted variables likely also bias upward the estimated endogenous social effect.

In Columns 4 and 5, rather than assuming that $\delta = 0$ (no contextual effects), I argue that there are a smaller number of variables that are validly excluded from the main equation. To be valid, these variables must meet several criteria: (1) the individual-level characteristic affects individual behavior (2) the school-level peer characteristic does not affect *individual* behavior, and (3) the school-level peer characteristic must affect *school-level* behavior.

I suggest that the number of older siblings and the presence of a household member who smokes meet these criteria. The intuition for the first exclusion restriction is that individuals with older siblings are more likely to smoke due to interactions with older individuals. However, the proportion of an individual's classmates who have older siblings is assumed to not affect the individual's smoking probability and only affect the proportion of an individual's classmates who smoke. Recall that I control for the total number of siblings, so the number of peers' older siblings should not be capturing differences in family sizes across schools. Similarly, I assume that while the presence of a household member who smokes increases an individual's propensity to smoke, the proportion of one's peers with a household member who smokes has no direct effect on the individual's smoking choices. Using these two instruments also allows the use of an over-identification test.

To provide further confidence in the instruments, I present results from balancing tests in Table AI in Appendix A, which shows that these instruments are essentially uncorrelated with individual- and family-level characteristics of the students (Lavy and Schlosser, 2007; Bifulco *et al.*, 2008).[24]

[22]I calculate the effect of moving from 0 to 25% prevalence in schools for comparative purposes with prior research. My own view is that examining small changes in peer smoking rates is likely more relevant for policy purposes.

[23]In Appendix Table AI, I also show that some of these instruments, such as grademates' average family income and maternal education, are correlated with several student-level characteristics such as test scores, maternal education, family income, and race. This is particularly true if school fixed effects are not controlled. These results suggest that students may be sorted based on the characteristics of their schoolmates and indicates that using these measures as instruments may be inappropriate. After controlling for school fixed effects, the correlations are much weaker, but a potential explanation for this result is that the amount of within-school variation for these grademate variables is small.

[24]Since 10 tests are run for each cohort-level instrument, we would expect to find approximately one statistically significant relationship by chance, which is what is shown.

DOI: 10.1002/hec

Table III. Individual-, family-, and peer-level determinants of adolescent smoking racial differences

Outcome	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS
Sample	White	Black	Hispanic
8th Grade	0.008	0.084	−0.124
	(0.041)	(0.051)	(0.118)
9th Grade	0.151***	−0.008	−0.082
	(0.041)	(0.042)	(0.070)
10th Grade	0.141***	−0.017	−0.054
	(0.048)	(0.046)	(0.077)
11th Grade	0.186***	−0.047	−0.049
	(0.051)	(0.055)	(0.091)
12th Grade	0.181***	−0.097	−0.132
	(0.060)	(0.064)	(0.089)
Age	0.003	0.026***	0.016
	(0.008)	(0.009)	(0.011)
Male	−0.028**	0.038**	0.044*
	(0.013)	(0.016)	(0.026)
PVT score	−0.004***	0.000	0.001
	(0.001)	(0.001)	(0.001)
Maternal education	0.002	−0.007*	0.001
	(0.003)	(0.003)	(0.004)
Family income	0.000	0.023	0.031
	(0.013)	(0.030)	(0.023)
Married household	−0.045***	−0.019	−0.005
	(0.014)	(0.012)	(0.016)
Household smoke	0.162***	0.043***	0.084***
	(0.013)	(0.014)	(0.021)
Religious attendance	−0.002***	−0.002***	−0.002***
	(0.000)	(0.000)	(0.000)
Number of older siblings	0.028***	0.016**	0.017***
	(0.007)	(0.006)	(0.005)
Number of siblings	−0.018***	−0.010**	−0.004
	(0.006)	(0.005)	(0.007)
Parent missing dummy	0.014	−0.003	−0.019
	(0.014)	(0.012)	(0.017)
Rural status	−0.004	0.001	0.002
	(0.014)	(0.023)	(0.039)
Urban status	−0.004	−0.009	0.019
	(0.017)	(0.017)	(0.018)
% Black	−0.006	−0.121***	0.012
	(0.066)	(0.031)	(0.114)
% Hispanic	−0.352***	−0.154***	−0.250***
	(0.079)	(0.035)	(0.056)
% Other race	0.257	−0.597*	0.119
	(0.327)	(0.315)	(0.280)
Mean family income	−0.001	−0.003***	0.001
	(0.001)	(0.001)	(0.002)
Mean maternal education	0.016	0.048***	−0.014
	(0.013)	(0.014)	(0.031)
% Married households	−0.001	−0.000	0.000
	(0.001)	(0.001)	(0.001)
Mean number of siblings	−0.012	0.004	0.041
	(0.022)	(0.019)	(0.026)
Mean religious attendance	0.001	−0.004***	−0.005*
	(0.001)	(0.001)	(0.003)
Number sampled	0.022	0.005	0.002
	(0.017)	(0.003)	(0.007)
Constant	0.402*	−0.530**	0.193
	(0.212)	(0.216)	(0.356)
Observations	6943	2748	2402
R-squared	0.09	0.05	0.06

Robust standard errors, ***1, **5, *10%.

DOI: 10.1002/hec

Table IV. Baseline and two-staged least squares results endogenous effects of peer smoking decisions on individual smoking outcomes

Outcome	Smoke	Smoke	% Smoke	Smoke	% Smoke
Specification	OLS	2SLS	First Stage	2SLS	First Stage
Sample	Full	Full	Full	Full	Full
Columns	1	2	3	4	5
Instruments		All contextual		%HH smoke Mean # older sibs	
% Smoke		0.481***		0.393***	
		(0.052)		(0.122)	
Grade	0.005	−0.003	0.018***	0.000	0.018***
	(0.008)	(0.006)	(0.005)	(0.006)	(0.005)
Age	0.014**	0.013**	0.003	0.013**	0.003
	(0.006)	(0.006)	(0.002)	(0.006)	(0.002)
Male	0.004	0.002	0.004**	0.002	0.004***
	(0.011)	(0.009)	(0.002)	(0.009)	(0.002)
Black	−0.191***	−0.170***	−0.003	−0.189***	−0.003
	(0.015)	(0.013)	(0.003)	(0.013)	(0.003)
Hispanic	−0.043**	−0.069***	−0.010**	−0.038**	−0.009**
	(0.017)	(0.016)	(0.004)	(0.017)	(0.004)
Other race	−0.058**	−0.064***	−0.003	−0.056***	−0.003
	(0.024)	(0.015)	(0.004)	(0.018)	(0.004)
PVT score	−0.001***	−0.001***	0.000	−0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal education	0.001	0.001	0.000	0.001	0.000
	(0.002)	(0.002)	(0.000)	(0.002)	(0.000)
Family income	0.003	0.007	−0.002	0.004	−0.002
	(0.009)	(0.009)	(0.001)	(0.009)	(0.001)
Married household	−0.033***	−0.031***	−0.001	−0.032***	−0.001
	(0.009)	(0.009)	(0.002)	(0.009)	(0.002)
Household smoke	0.127***	0.126***	0.001	0.125***	0.001
	(0.013)	(0.009)	(0.001)	(0.009)	(0.001)
Religious attendance	−0.002***	−0.002***	0.000	−0.002***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of older siblings	0.022***	0.021***	0.001	0.021***	0.001
	(0.005)	(0.004)	(0.001)	(0.004)	(0.001)
Number of siblings	−0.014***	−0.013***	−0.000	−0.013***	0.000
	(0.004)	(0.004)	(0.001)	(0.004)	(0.001)
Parent missing dummy	0.002	0.006	−0.003	0.003	−0.003
	(0.009)	(0.009)	(0.002)	(0.009)	(0.002)
Rural status	0.002	−0.001	0.011*	−0.003	0.011*
	(0.010)	(0.010)	(0.007)	(0.011)	(0.007)
Urban status	0.002	−0.000	−0.001	0.002	−0.001
	(0.010)	(0.008)	(0.004)	(0.008)	(0.004)
% Male	0.002***		0.002***	0.001*	0.002***
	(0.001)		(0.001)	(0.000)	(0.001)
% Black	−0.088**		−0.243***	0.024	−0.254***
	(0.038)		(0.037)	(0.045)	(0.036)
% Hispanic	−0.254***		−0.242***	−0.122***	−0.248***
	(0.044)		(0.038)	(0.045)	(0.038)
% Other race	−0.094		−0.154***	−0.022	−0.161***
	(0.058)		(0.047)	(0.042)	(0.047)
Mean family income	−0.001*		−0.001	−0.001	−0.001
	(0.001)		(0.001)	(0.000)	(0.001)
Mean maternal education	0.016		0.011	0.017**	0.011
	(0.012)		(0.011)	(0.008)	(0.011)
% Married households	−0.000		−0.000	−0.000	−0.000
	(0.001)		(0.000)	(0.000)	(0.000)
Mean number of siblings	0.003		0.004	−0.000	−0.027
	(0.018)		(0.017)	(0.010)	(0.024)
Mean religious attendance	−0.001		−0.003***	0.000	−0.003***
	(0.001)		(0.001)	(0.001)	(0.001)

DOI: 10.1002/hec

Table IV. *Continued*

Outcome	Smoke	Smoke	% Smoke	Smoke	% Smoke
Specification	OLS	2SLS	First Stage	2SLS	First Stage
Sample	Full	Full	Full	Full	Full
Columns	1	2	3	4	5
Number sampled	0.011**	0.002	0.010**	0.006**	0.011**
	(0.005)	(0.002)	(0.005)	(0.003)	(0.004)
% Household smokes			0.221***		0.212***
			(0.049)		(0.050)
Mean number older siblings			0.114**		0.083***
			(0.049)		(0.031)
Constant	0.017	0.090	−0.153	−0.111	−0.119
	(0.149)	(0.065)	(0.155)	(0.107)	(0.155)
Observations	13 316	13 316	13 316	13 316	13 316
R-squared	0.10	0.10	0.59	0.10	0.60
J-statistic *p*-value		0.000		0.280	
F-statistic			30.93		13.92

Robust standard errors, ***1, **5, *10%.

Importantly, the instruments are shown to be uncorrelated with the student's performance on a standardized test, the student's family income, and maternal education and other characteristics. This is suggestive evidence that these instruments can be thought of as plausibly exogenous within school and that parents are not systematically changing student's schools based on these cohort characteristics.

The empirical results using the preferred set of instruments are shown in Columns 4 (main results) and 5 (first stage results). Interestingly, the coefficient representing an endogenous social effect decreases to 0.39, indicating that moving a student from a school with 0% smoking to 25% smoking increases the individual's likelihood of smoking by less than 10 percentage points. The *F*-statistic for this specification is lower (13.92) but still greater than the cutoff of 10 suggested by Stock and Staiger (1997). Finally, I fail to reject the validity of the instruments at the 28% level. As indicated above, it is important to note that these specifications in Table IV have not controlled for group-level unobservable factors that could simultaneously affect individual and peer-group outcomes or the endogeneity of peer outcomes due to residential sorting, likely upwardly biasing the estimated endogenous effect coefficients. This omission is standard in much previous research on social interactions. I turn to this issue in Table V.

Table V presents results from specifications that employ a school-level fixed effects/instrumental variables strategy. Like the results of Table IV, the instrumental variables are used to assure identification of the endogenous effect, but unlike Table IV, omitted characteristics of the schools, which could simultaneously affect individual- and group-level choices, are controlled through a school-level fixed effect. It is important to examine the extent of within-school variation in peer-level variables when using the school-level fixed effects. In the table in Appendix A, I provide evidence of the variation within schools in the grade-level characteristics. Importantly, there is substantial variation in the grade-level proportion smoking, mean number of older siblings, and proportion of household members who smoke. There is less variation in the racial composition of grades within a school.

In Columns 1 and 2 in Table V, I present results that again assume that $\delta = 0$ and that all peer-level characteristics are valid instruments for peer-level smoking choices. With school-level fixed effects, the endogenous effect suggest that moving a student from a school with 0% smoking to 25% smoking increases the individual's likelihood of smoking by over 12 percentage points, which is nearly identical to the results without fixed effects that use the questionable set of instruments. Unfortunately, including school-level fixed effects reduces the *F*-statistic to under 4 because there is little variation within schools with several of the instruments (% black, etc.).

DOI: 10.1002/hec

Table V. Two-staged least squares results with school-level fixed effects endogenous effects of peer smoking decisions on individual smoking outcomes

Outcome	Smoke	% Smoke	Smoke	% Smoke
Specification	2SLS/FE	First stage	2SLS/FE	First stage
Instruments	All contextual		% HH smoke Mean # older sibs	
Sample	Full	Full	Full	Full
Columns	1	2	3	4
% Smoke	0.488***		0.345**	
	(0.113)		(0.171)	
Grade	−0.001	0.022***	0.003	0.022***
	(0.007)	(0.005)	(0.008)	(0.005)
Age	0.013**	0.003**	0.014**	0.003**
	(0.006)	(0.001)	(0.006)	(0.001)
Male	0.000	0.001	0.001	0.001
	(0.012)	(0.002)	(0.012)	(0.002)
Black	−0.192***	−0.003	−0.193***	−0.003
	(0.015)	(0.002)	(0.015)	(0.002)
Hispanic	−0.037**	−0.005*	−0.037**	−0.005*
	(0.016)	(0.003)	(0.016)	(0.003)
Other race	−0.059**	−0.002	−0.060**	−0.002
	(0.025)	(0.002)	(0.025)	(0.002)
PVT score	−0.001***	0.000	−0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Maternal education	0.001	−0.000	0.001	−0.000
	(0.002)	(0.000)	(0.002)	(0.000)
Family income	0.005	0.000	0.005	0.000
	(0.009)	(0.002)	(0.009)	(0.002)
Married household	−0.031***	0.000	−0.031***	0.000
	(0.009)	(0.002)	(0.009)	(0.002)
Household smoke	0.125***	0.000	0.125***	0.000
	(0.013)	(0.002)	(0.013)	(0.002)
Religious attendance	−0.002***	0.000	−0.002***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of older siblings	0.021***	0.001	0.022***	0.001
	(0.004)	(0.001)	(0.004)	(0.001)
Number of siblings	−0.013***	−0.001	−0.013***	−0.001
	(0.004)	(0.001)	(0.004)	(0.001)
Parent missing dummy	0.005	−0.001	0.005	−0.001
	(0.009)	(0.002)	(0.009)	(0.002)
Rural status	−0.009	0.004**	−0.008	0.004**
	(0.011)	(0.002)	(0.011)	(0.002)
Urban status	0.003	0.003	0.004	0.003
	(0.010)	(0.002)	(0.010)	(0.002)
% Male		0.001	0.000	0.001
		(0.001)	(0.000)	(0.001)
% Black		−0.376***	−0.096	−0.376***
		(0.120)	(0.085)	(0.120)
% Hispanic		0.217	0.257**	0.217
		(0.141)	(0.110)	(0.141)
% Other race		−0.132	−0.029	−0.132
		(0.107)	(0.077)	(0.107)
Mean family income		0.000	−0.000	0.000
		(0.001)	(0.001)	(0.001)
Mean maternal education		0.001	0.009	0.001
		(0.015)	(0.011)	(0.015)
% Married households		−0.000	0.000	−0.000
		(0.001)	(0.000)	(0.001)
Mean number of siblings		−0.059	0.000	−0.059
		(0.038)	(0.018)	(0.038)
Mean religious attendance		−0.002	0.000	−0.002
		(0.002)	(0.001)	(0.002)

DOI: 10.1002/hec

Table V. *Continued*

Outcome	Smoke	% Smoke	Smoke	% Smoke
Specification	2SLS/FE	First stage	2SLS/FE	First stage
Instruments	All contextual		% HH smoke Mean # older sibs	
Sample	Full	Full	Full	Full
Columns	1	2	3	4
Number sampled	0.037*	0.055**	0.042*	0.055**
	(0.020)	(0.027)	(0.023)	(0.027)
% Household smokes		0.102***		0.102***
		(0.036)		(0.036)
Mean number older siblings		0.163**		0.163**
		(0.076)		(0.076)
Constant		−0.118		−0.118
		(0.201)		(0.201)
Observations	13 316	13 316	13 316	13 316
R-squared	0.05	0.21	0.06	0.21
J-statistic *p*-value	0.260		0.130	
Number of schools	72	72	72	72
F-statistic		3.444		7.443

Robust standard errors, ***1, **5, *10%.

In Columns 3–4, I repeat the results from Table IV but include school-level fixed effects with the limited set of instruments. As expected, the estimated endogenous effect decreases in magnitude after using school fixed effects from 0.39 to 0.34. The endogenous effect suggests that moving a student from a school with 0% smoking to 25% smoking increases the individual's likelihood of smoking by 8 percentage points, which decreases the magnitude of the results without fixed effects by over 10% and are 30% lower than the results that assume no contextual effects, an assumption that is normally made in the literature. The results have a moderate *F*-statistic (7.4) and also comfortably fail to reject the validity of the instruments used (13%). Since the *F*-statistic is moderate, we may be somewhat concerned that the estimated peer effects may still be slightly biased upward.[25] The reader should keep this issue in mind when interpreting the results.

5.1. Falsification tests

Finally, in Appendix Table AII, I follow the procedure outlined in Lavy and Schlosser (2007) to implement falsification tests. In particular, I replace the observed proportion of peers who smoke with the measure of peer smokers at a lower or higher grade within the same school. The procedure is used to provide additional evidence that the results of the preferred specifications are not biased from an omitted unobservable confounder of the coefficient of interest. Importantly, I find no endogenous peer effect when using the 'placebo treatments'.

6. CONCLUSIONS

In this paper, I use a social interactions framework to detect whether individual smoking decisions are influenced by classmate smoking decisions. I address several large challenges in addressing this question,

[25]I attempted to construct test-statistics that take into account potentially weak instruments, such as the Stata command 'condivreg' (Mikusheva and Poi, 2006). Unfortunately, current statistical methods in Stata do not allow the use of fixed effects within these tests. The results from Table IV, which do not use school fixed effects, are robust to these corrections for weak instruments, though.

DOI: 10.1002/hec

including the endogeneity of school (and thus classmates) through residential location choices, 'third factors' such as school-level unobservables that influence individual and classmate choices simultaneously, and the difficulty of the identification of parameters in empirical models of social interactions (Manski, 1993; Brock and Durlauf 2001). In particular, I use an instrumental variables/fixed effects methodology that compares students in different grades within the same high school who face a different set of classmates and classmates' decisions. I provide evidence that the instruments are plausibly exogenous and also pass the over-identification tests. Preferred specifications suggest that increasing the proportion of classmates who smoke by 10% will increase the likelihood an individual smokes by approximately 3 percentage points. Falsification tests support the validity of the results. I also compare these results with previous findings that are unable to use school fixed effects and/or use potentially invalid instruments and find evidence that using the preferred methodology suggests smaller peer effects.

While this paper is able to address many of the relevant econometric issues in estimating the importance of social influences on smoking decisions, there are several limitations. First, while I follow much of the current literature in assuming that the relevant reference group for adolescents are classmates, it could be the case that this reference group is too broad, too narrow, or of limited importance in influencing the decisions for typical adolescents. A second limitation is that there could be school-grade specific correlated effects that are not eliminated through using school fixed effects.

Even with the limitations of the study, I find relatively strong and consistent evidence that peer smoking choices provide a moderate influence on adolescents' decisions to smoke. These findings suggest that policies that influence one person's decision to smoke will also affect the decisions of the person's classmates through social interactions. These results suggest that traditional cost–effectiveness analysis of smoking intervention programs that do not take into account the spillovers effect on untreated individuals likely understate the total social benefits of the programs and suggest that too few programs will be undertaken. Finally, I find evidence that previous research that has failed to use group-level fixed effects or have assumed no contextual effects and pursued an instrumental variables strategy with peer-level characteristics likely produce incorrect estimates of endogenous social effects of smoking.

APPENDIX A

Variation in grade-level characteristics controlling for school and grade fixed effects.

	R-squared Xs: school and grade dummies
% Smoke	0.72
Mean number of older sibs	0.62
% Household smoking	0.80
% Black	0.97
% Hispanic	0.97
% Other race	0.78
% Married	0.78
% Male	0.50

Results from balancing tests are given in Table AI.

The procedure outlined in Lavy and Schlosser (2007) to implement falsification tests is given in Table AII.

DOI: 10.1002/hec

Table AI. Balancing tests for the association between the (A) cohort-level instruments and individual characteristics and (B) potentially invalid instruments and individual characteristics

Outcome	Male	Black	Hispanic	PVT score	Maternal education	Family income	Married parents	Religiosity	Older siblings	Number of siblings
Specification	OLS									
(A)										
Mean HH smoking	−0.052	−0.019	−0.035	−0.435	−0.161	0.019	−0.081	1.597	−0.148	−0.364*
	(0.054)	(0.044)	(0.036)	(2.011)	(0.279)	(0.050)	(0.061)	(3.178)	(0.184)	(0.194)
Observations	14 492	14 492	14 492	13 800	14 492	14 492	14 492	14 475	14 473	14 292
F-test	0.929	0.194	0.902	0.047	0.333	0.147	1.769	0.253	0.650	3.514
F-*p*-value	0.338	0.661	0.346	0.829	0.566	0.703	0.188	0.617	0.423	0.0650
Mean number older sib	0.019	0.014	0.092**	−2.321	0.006	−0.035	0.031	0.451	−0.229	−0.183
	(0.040)	(0.038)	(0.038)	(1.809)	(0.244)	(0.053)	(0.048)	(2.656)	(0.307)	(0.300)
Observations	14 473	14 473	14 473	13 782	14 473	14 473	14 473	14 457	14 473	14 275
F-test	0.233	0.137	5.940	1.646	0.001	0.443	0.412	0.029	0.558	0.372
F-*p*-value	0.631	0.713	0.0173	0.204	0.980	0.508	0.523	0.866	0.457	0.544

(B)

Outcome	Black	Black	Hispanic	Hispanic	PVT score	PVT score	Maternal education	Maternal education	Married	Married	Household smoke	Household smoke	Number sibs	Number sibs
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed effects	Grade	School and grade	Grade	School and grade	Grade	School and grade	Grade	School and grade	Grade	School and grade	Grade	School and grade	Grade	School and grade
Mean family income	−0.004**	−0.000	−0.005**	−0.000	0.280***	0.027	0.040***	0.002	0.003***	0.000	−0.002***	0.000	−0.006**	−0.002
	(0.002)	(0.001)	(0.002)	(0.000)	(0.030)	(0.029)	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Observations	14 492	14 492	14 492	14 492	13 800	13 800	14 492	14 492	14 492	14 492	14 492	14 492	14 292	14 292
R-squared	0.033	0.359	0.046	0.364	0.103	0.198	0.083	0.137	0.016	0.053	0.007	0.063	0.005	0.053
F-test	5.604	0.443	4.493	0.323	88.172	0.851	141.224	0.260	20.120	0.038	10.671	0.138	5.985	1.143
F-*p*-value	0.0207	0.508	0.0375	0.572	0	0.359	0	0.611	0.0000274	0.847	0.00168	0.711	0.0169	0.289
Mean maternal educ	0.026	−0.015	−0.138***	0.001	4.905***	0.147	0.137***	0.009	0.039***	0.030**	−0.048**	−0.007	−0.110	0.105**
	(0.026)	(0.015)	(0.046)	(0.008)	(0.549)	(0.636)	(0.021)	(0.013)	(0.011)	(0.014)	(0.018)	(0.014)	(0.068)	(0.051)
Observations	14 492	14 492	14 492	14 492	13 800	13 800	14 492	14 492	14 492	14 492	14 492	14 492	14 292	14 292
R-squared	0.005	0.359	0.108	0.364	0.091	0.198	0.084	0.138	0.008	0.053	0.009	0.063	0.005	0.053
F-test	1.021	0.974	9.079	0.017	79.925	0.053	43.488	0.456	13.360	4.380	6.961	0.262	2.597	4.285
F-*p*-value	0.316	0.327	0.00358	0.896	0	0.818	6.45e−09	0.502	0.000489	0.0399	0.0102	0.611	0.111	0.0421

Additional controls: complete set of school dummies and grade dummies. Regressions are weighted and standard errors are allowed to be clustered at the school level. Robust standard errors, ***1, **5, *10%.

DOI: 10.1002/hec

Table AII. Falsification tests. Replace grade-level proportion smoking with earlier grade within school

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Specification	2SLS/FE	First stage/FE	2SLS/FE	First stage/FE	2SLS/FE	First stage/FE
Sample	Full	Full	Full	Full	Full	Full
% Smoke	0.345**		−0.898		−1.779	
	(0.176)		(0.894)		(1.792)	
Grade	0.003	0.021***	0.042	0.028***	0.056	0.023***
	(0.008)	(0.005)	(0.030)	(0.007)	(0.052)	(0.007)
Age	0.014**	0.003**	0.009	0.001	0.022***	0.002*
	(0.006)	(0.001)	(0.006)	(0.001)	(0.008)	(0.001)
Male	0.000	0.000	0.010	−0.000	−0.015	−0.001
	(0.012)	(0.001)	(0.011)	(0.001)	(0.012)	(0.001)
Black	−0.192***	−0.003	−0.191***	−0.000	−0.185***	0.001
	(0.015)	(0.002)	(0.017)	(0.002)	(0.014)	(0.002)
Hispanic	−0.037**	−0.005*	−0.022	0.000	−0.022	0.001
	(0.016)	(0.002)	(0.019)	(0.002)	(0.019)	(0.002)
Other race	−0.060**	−0.002	−0.061**	0.001	−0.074***	−0.002
	(0.025)	(0.002)	(0.027)	(0.003)	(0.027)	(0.002)
PVT score	−0.001***	0.000	−0.001***	−0.000	−0.001***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal education	0.001	−0.000	0.003	0.000	0.001	−0.000
	(0.002)	(0.000)	(0.002)	(0.000)	(0.002)	(0.000)
Family income	0.005	0.000	0.006	0.000	0.004	0.001
	(0.009)	(0.002)	(0.009)	(0.002)	(0.011)	(0.002)
Married household	−0.031***	0.000	−0.026**	−0.001	−0.032***	0.001
	(0.009)	(0.002)	(0.011)	(0.002)	(0.010)	(0.002)
Household smoke	0.125***	0.000	0.130***	0.003	0.123***	0.000
	(0.013)	(0.002)	(0.014)	(0.002)	(0.014)	(0.002)
Religious attendance	−0.002***	0.000	−0.002***	−0.000	−0.002***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Number of older siblings	0.022***	0.001	0.020***	−0.001	0.024***	−0.001
	(0.004)	(0.001)	(0.005)	(0.001)	(0.005)	(0.001)
Number of siblings	−0.013***	−0.001	−0.013***	0.001	−0.017***	0.000
	(0.004)	(0.001)	(0.005)	(0.001)	(0.004)	(0.001)
Parent missing dummy	0.005	−0.001	0.017	0.001	−0.003	−0.000
	(0.009)	(0.002)	(0.011)	(0.002)	(0.010)	(0.002)
Rural status	−0.009	0.004**	−0.002	−0.000	0.002	0.003*
	(0.010)	(0.002)	(0.014)	(0.002)	(0.014)	(0.002)
Urban status		0.003*		0.002		0.001
		(0.002)		(0.002)		(0.002)
% Black	−0.094	−0.370***	−0.372	0.231**	−0.675	−0.267*
	(0.087)	(0.123)	(0.265)	(0.108)	(0.580)	(0.153)
% Hispanic	0.256**	0.217	0.350	−0.185	0.879	0.293
	(0.111)	(0.140)	(0.274)	(0.233)	(0.581)	(0.199)
% Other race	−0.026	−0.125	−0.062	−0.047	−0.461	−0.159
	(0.079)	(0.106)	(0.284)	(0.180)	(0.528)	(0.176)
Mean family income	−0.000	0.000	−0.001	0.000	−0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Mean maternal education	0.009	0.001	0.029	0.010	−0.020	−0.010
	(0.011)	(0.015)	(0.024)	(0.018)	(0.037)	(0.019)
% Married households	0.000	−0.000	0.000	−0.001	0.001	0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Mean religious attendance	0.000	−0.002	−0.003	−0.003	0.003	0.002
	(0.001)	(0.002)	(0.004)	(0.002)	(0.004)	(0.002)
Number sampled	0.042*	0.055**	0.103**	0.039**	0.062	0.074
	(0.023)	(0.027)	(0.051)	(0.019)	(0.194)	(0.084)
Mean number of slblings	0.000	−0.060	0.013	0.041	−0.055	0.001
	(0.018)	(0.038)	(0.036)	(0.043)	(0.078)	(0.043)
Number of older siblings		0.101***		−0.032		−0.044
		(0.036)		(0.051)		(0.043)
% Households smoke		0.157**		0.100		0.035
		(0.076)		(0.088)		(0.083)
Constant		−0.066		−0.112		0.005
		(0.200)		(0.236)		(0.296)

DOI: 10.1002/hec

Table AII. *Continued*

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Specification	2SLS/FE	First stage/FE	2SLS/FE	First stage/FE	2SLS/FE	First stage/FE
Sample	Full	Full	Full	Full	Full	Full
Observations	13 316	13 316	9827	9883	10 447	10 505
R-squared	0.06	0.21	0.05	0.21	0.01	0.14
Number of schools	72	72	72	72	72	72
J-statistic *p*-value	0.127		0.431		0.675	
F-statistic		7.309		0.764		0.542

Note: Columns 1 and 2 present baseline results. Columns 3 and 4 show results of the second placebo treatment, when the cohorts are rotated backward (11th graders' %smoking replaced by 10th graders') and 7th graders are dropped in placebo ♯1. Since this calculation is done within schools, individuals in schools without a lower grade (e.g. 9th grades in high schools) are also dropped. Columns 5 and 6 show results of the first placebo treatment, when the cohorts are rotated forward (11th graders' %smoking replaced by 12th graders') and 12th graders are dropped in placebo ♯2. Robust standard errors, ***1, **5, *10%.

ACKNOWLEDGEMENTS

This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). I thank Ethan Cohen-Cole, Philip DeCicca, Andrew Grodner, Robert Kaestner, Steven Lehrer, John Mullahy, Gabriel Picone, Steve Pischke, Analia Schlosser and participants at East Carolina University, University of Illinois-Chicago, and the 2007 Southeastern Health Economics Study Group meeting for helpful comments. All errors are my own.

REFERENCES

Bifulco R, Fletcher JM, Ross SL. 2008. The effect of classmate characteristics on individual outcomes: evidence from the Add Health. *SSRN Working Paper*. (Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id = 1259570.)

Brock W, Durlauf S. 2001 Discrete choice with social interactions. *Review of Economic Studies* **68**(2): 235–260.

Case A, Katz L. 1991. The company you keep: the effects of family and neighborhood on disadvantaged youth. *NBER Working Paper 3705*.

Chaloupka FJ, Wechsler H. 1997. Price, tobacco control policies and smoking among young adults. *Journal of Health Economics* **16**(3): 359–373.

Clark A, Loheac Y. 2007. It wasn't me, it was them! social influence in risky behavior by adolescents. *Journal of Health Economics* **26**(4): 763–784.

DeCicca P, Kenkle D, Mathios A. 2002. Putting out the fires: will higher taxes reduce the onset of youth smoking? *Journal of Political Economy* **110**(1): 144–169.

DeCicca P, Kenkle D, Mathios A, Shin Y-J, Lim J-Y. 2006 Youth smoking, cigarette prices, and anti-smoking sentiment. *NBER Working Paper 12458*.

Douglas S. 1998. The duration of the smoking habit. *Economic Inquiry* **36**(1): 49–64.

Douglas S, Hariharan G. 1994. The hazard of starting smoking: estimates from a split population duration model. *Journal of Health Economics* **18**:429–441.

Fletcher JM. 2006. Social interaction and college enrollment: evidence from NELS. *Working Paper*.

Fletcher JM. 2007. Social multipliers in the sexual initiation decisions among U.S. high school students. *Demography* **44**(2): 373–388.

Fletcher JM. 2008. Peer influences on alcohol consumption. *Working Paper*, Yale University.

DOI: 10.1002/hec

Fletcher JM, Tienda M. 2008. Peer networks and college success: lessons from Texas. *Working Paper*, Yale University.
Gaviria A, Raphael S. 2001. School-based peer effects and juvenile behavior. *Review of Economics and Statistics* **83**(2): 257–268.
Hanushek EA, Kain JF, Markman JM, Rivkin SG. 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics* **18**(5): 527–544.
Hoxby C. 1999. The effects of school choice on curriculum and atmosphere. In *Earning and Learning*: *how Schools Matter*, Mayer S, Peterson P (eds). Brookings Institute Press: New York.
Hoxby C. 2000. Peer effects in the classroom: learning from gender and race variation. *NBER Working Paper 7867*.
Kawaguchi D. 2004. Peer effects on substance use among American teenagers. *Journal of Population Economics* **17**: 351–367.
Kremer M, Levy V. 2003. Peer effects and alcohol use among college students. *NBER Working Paper 9876*.
Lavy V, Schlosser A. 2007. Mechanisms and impacts of gender peer effects at school. *Working Paper*.
Luccassen R, Andrew R, Coats M, Karahan G. 2005. Cigarette smuggling mitigates the public health benefits of cigarette taxes. *Applied Economics Letters* **12**(12): 769–773.
Lundborg P. 2006. Having the wrong friends? *Peer effects in adolescent substance use*. *Journal of Health Economics* **25**: 214–233.
Manski C. 1993. Identification of endogenous social effects: the reflection problem. *Review of Economic Studies* **60**(3): 531–542.
Manski C. 1995. *Identification Problems in the Social Sciences*. Harvard University Press: Cambridge, MA.
Manski C. 2000. Economic analysis of social interactions. *Journal of Economic Perspectives*. **14**: 3.
Mikusheva A, Poi B. 2006. Tests and confidence sets with correct size when instruments are potentially weak. *Stata Journal* **6**(3): 335–347.
Powell L, Taurus J, Ross H. 2005. The importance of peer effects, cigarette prices, and tobacco control policies for youth smoking behavior. *Journal of Health Economics* **24**: 950–968.
Sacerdote B. 2001. Peer effects with random assignments: results from Dartmouth roommates. *Quarterly Journal of Economics* **116**: 2.
Summers A, Wolfe B. 1977. Do schools make a difference? *American Economic Review* **67**(4): 639–652.
Vigdor J, Nechyba T. 2004. Peer effects in elementary school: learning from apparent random assignment. *Working Paper*.

DOI: 10.1002/hec