Estimating the Effects of Friendship Networks on Health Behaviors of Adolescents[*]

Jason M. Fletcher
Yale University

Stephen L. Ross
University of Connecticut

This Draft: February 6, 2010

Abstract

Researchers typically examine peer effects by defining the peer group broadly (all classmates, schoolmates, neighbors) because of the lack of friendship information in many data sources as well as to enable the use of plausibly exogenous variation in peer group composition across cohorts in the same school. This paper estimates the effects of friend's health behaviors on own health behaviors for adolescents. A causal effect of friend's health behaviors is identified by comparing similar individuals who have the same friendship opportunities because they attend the same school and make the same friendship choices, under the assumption that the friendship choice reveals information about an individual's unobservables. We combine this identification strategy with a cross-cohort, within school design so that the model is identified based on across grade differences in the clustering of health behaviors within specific friendship options. This strategy allows us to separate the effect of friends behavior on own behavior from the effect of friends observables attributes on behavior, a key aspect of the reflection problem. Our results suggest that friendship network effects are important in determining adolescent tobacco and alcohol use but are over-estimated in specifications that do not fully take into account the endogeneity of friendship selection by 20-30%

[*] We received valuable comments from seminar participants at Baylor University, Lafayette College, Lehigh University, Texas A&M, University of Texas-Austin, and Yale University.
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 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).

Introduction

Individuals in modern societies are socially connected in a multitude of ways. For example, the social networking website Facebook.com has increased its membership by 100 million users in the last eight months (as of July 2009), bringing the total to over 200 million users worldwide. Individuals use their social networks to receive and send information as well as establish, update, and enforce social norms of behavior. Both information acquisition as well as the impacts of social norms within social networks could have large effects on the health behaviors of individuals, particularly adolescents, who are particularly responsive to peer pressure (Brown et al. 1997). This heightening of peer influence also takes place during the developmental stage when many of the most costly health outcomes and behaviors are initiated.

This confluence of events during adolescence sets the stage for a critical period of individual decision-making with long-term consequences and also provides a unique window of opportunity for health intervention. While much recent research has shown robust and important correlations in the health behaviors of socially connected individuals, whether the correlations are causal is still largely unknown for many health behaviors. Unlike research on broader peer groups, most work on the effect of friendships and/or the associated networks are identified using information that arises from connections between individuals within broader peer groups assuming that those connections or links are exogenous. .

In this paper, we implement a new method of estimating the causal effects of friendship networks. Following logic similar to Dale and Krueger (2002), we assume that observationally similar individuals who face the same friendship opportunity set must be similar on unobservables if they make similar friendship choices. Accordingly, we build on the large literature suggesting that individuals exhibit strong racial, gender, and age preferences when choosing their friends—likes choose likes (Mayer and Puller 2008, Weinberg 2008), and test whether friend's behavior influences own behavior among groups of similar individuals who chose observationally similar sets of friends. Specifically, we use unique data from the Add Health, which contains information both on friendship nominations and health behaviors for a national sample of adolescents, and control for friendship *options* and friendship *choices* using school and friendship choice

pattern fixed effects, respectively. The inclusion of fixed effects for friendship choices yield consistent estimates of the effect of behavior under the assumption that behavior differences across similar friendship groups for individuals in the same cluster are unrelated to individual unobservables that influence the behavior. Most significantly, this assumption allows us to separate the influence on individual behavior of friend's behaviors from the influence of the observable attributes of those friends because comparisons are made between individuals who have observationally equivalent sets of friends.

We then use a feature of the Add Health that multiple cohorts of students are surveyed in each school in order to enhance our identification strategy. We identify the effect of peer behaviors by comparing individuals in different cohorts at the same school who make similar friendship choices. The assumption that unobservables are uncorrelated with behavior differences across friendship groups is more plausible because, given the strong pattern of within grade friendships, an individual did not have the comparison group's friends as options when forming or reporting their list of friends. This strategy can be illustrated by the following thought exercise: consider a 9^{th} grader and 10^{th} grader who attend the same high school. As we show in detail below, these students face very similar friendship opportunities with respect to racial, gender, and socioeconomic composition of their same-grade classmates, and yet there is substantial clustering of health behaviors into specific cohorts within schools. Thus, if we compare two students who choose similar "types" of friends based on race, maternal education, and other demographic characteristics, there will exist substantial differences in health behaviors between the across cohort friendship opportunities, and those differences in friends' health behaviors is arguably quasi-random. The key is that the age difference between the 9^{th} grader and the 10^{th} grader (who attend the same high school and have the same preferences for "types" of friends) has effectively randomized these two students into their actual friendship network.

We find evidence that this strategy produces smaller "network effect" estimates than more standard models; however we still find robust evidence of network effects on smoking and drinking behavior of adolescents. Further, we find that peer health behaviors are statistically insignificant predictors of predetermined student or family

attributes and of health outcomes that should arguably be unaffected by social influences like chest pain, and the estimated coefficients in these models are much smaller than our estimates of the effect on health behaviors.

Background Literature

A large body of research across multiple disciplines has shown very strong correlations in health behaviors for individuals who are socially connected. One reason there has been so much research and policy interest in exploring how networks affect health behaviors and outcomes is the potentially large set of health interventions and policies that could be proposed to leverage social influences on health behaviors. While the promise of using social networks to affect health is compelling, so too are the empirical issues inherent in detecting causal effects of social networks using observational data.

Four difficulties with estimating the causal effects of social networks on health are particularly important (Manski 1993). First, individuals self-select into their social network; smokers befriend smokers. Second, individuals in the same social network are simultaneously affected by their shared environment; common exposure to a smoking ban likely reduces tobacco use among all members of a social network. Third, it is difficult to separate the influence of an individual's behavior and an individual's attributes in determining the health behaviors of his or her friend. Fourth, social influences are likely reciprocal, which leads to simultaneity bias. Unfortunately, failure to overcome these empirical difficulties casts considerable doubt on the current knowledge base linking the health behaviors among individuals in the same social network. Each of these biases can lead a researcher to incorrectly infer that social networks have a causal influence on behavior. Thus, policies intended to utilize social networks to enhance interventions to reduce unhealthy behaviors could be unable to affect change if social networks do not actually have causal effects. Providing evidence of the causal mechanisms and the likely effects of policies is essential to be able to properly leverage social network effects on health behaviors.

There have been two directions that researchers have taken in estimating peer effects on health behaviors: [1] focus on broadly defined peer groups, such as all

classmates in a school, in order to either (a) exploit cross-cohort population variation[1] in classmate composition (Bifulco et al. 2008, Fletcher in press, 2008, Trogdon et al. 2008, Lundburg 2006, Clark and Loheac 2007) and/or (b) use instrumental variable strategies (Powell et al. 2005, Gaviria and Raphael 2001[2]) or [2] focus on narrowly defined peer groups, such as nominated friends, where the issues with endogeneity are thornier and the estimates are likely less credible (Troddon et al. 2008, Christakis and Fowler 2007, 2008, Renna et al. 2008). In this paper, we seek to combine the more credible research designs from the first literature with the more credible peer group definitions of the second literature.

Since we focus on friendship networks as the definition of peer group in this paper, it is necessary to outline what other researchers have done previously and how our strategy adds to the literature in this area. There have been recent examinations of the effects of social networks on obesity and smoking in the medical literature (Christakis and Fowler 2007, 2008), where "friends" are measured by the names respondents provide as potential contact sources for future survey waves. In order to control for endogeneity of friendships, Christakis and Fowler assume that including lags of the outcome for both the respondent and his/her friend is sufficient. They also do not control for common environmental factors. Cohen-Cole and Fletcher (2008a) show that adding controls for environmental factors eliminates any detectable social network effects for obesity, and Cohen-Cole and Fletcher (2008b) show more generally that these parsimonious models will produce social network effects even in outcomes where none exist, such as for height.

Renna et al. (2008) and Trodgon et al. (2008) also focus on estimating social contagion in obesity and control for endogeneity of friendship in part by using school fixed effects. However, since friendship sorting occurs both across schools and within schools, school fixed effects likely do not provide a full solution to the endogeneity of friend selection, unless students select friends randomly within schools. Renna et al. (2008) and Trodgon et al. (2008) use instruments for friends' weight, including friends'

[1] See also the similar literature estimating peer effects in education outcomes (Hoxby 2000, Lavy and Schlosser 2008, Hanushek et al. 2003)

[2] Instruments used in these analyses are often questionable, such as census poverty measures. Fletcher (in press) provides suggestive evidence that these instruments are invalid and proposes alternatives. Trodgon et al. (2008) and Fletcher (in press) use a combination of fixed effects and instruments.

parents' obesity. Trodgon et al. also uses friends' birth weight. It is unclear whether these instruments are adequate, though, as they are known at the time of friendship selection.

Calvó-Armengol et al. (in press) have extended the literature by using a network fixed effects approach in their examination of peer effect in education outcomes. Adolescents are assumed to choose among mutually exclusive networks of friends. Within these networks, their best friends (based on friendship nominations) are used as the peer exposure and network fixed effects are controlled. The assumption with this approach is that adolescents endogenously choose a friendship group, but within that group, actual "best friends" are random.

In this paper, we adapt and extend a recent estimation method developed by Dale and Krueger (2002). In the original application, Dale and Krueger estimate the effects of exposure to "high ability" peers in college on an individual's own performance during college. As outlined above, a critical empirical difficulty is modeling the fact that college selection (and thus peer selection) is not random, but instead a choice of the student. Dale and Krueger use a large survey of college students that contains information on the submitted applications as well as the application decisions (options and choices) of the students and colleges. With this information the authors are able to compare the college outcomes of students who applied to and were accepted into the same types of colleges. As these students are likely quite similar in both observable characteristics such as SAT scores as well as unobservable characteristics such as motivation, a comparison between two students with similar college application and college acceptance patterns should reduce the selection bias in the estimates. Importantly, in this particular application, Dale and Krueger found that the "social network effect" of attending a college with high ability peers that is typically estimated in models that do not adequately control for selection into peer groups was completely eliminated using this procedure.

We will adapt the intuition behind this approach to focus on friendship networks and the potential influence of friend's health behavior on a student's own health behavior. Specifically, conditioning on school and an individuals observable attributes assures that we are comparing observationally similar individuals facing similar friendship opportunities, and if those similar individuals make the same friendship choices then it is

unlikely that there are substantial differences between them on unobservables. Further, we extend the methodology to include a more credible source of exogenous variation by including a cross-cohort research design.

Friendship Data

In order to accomplish our research goals, we use the only available national dataset containing rich friendship network information as well as health behaviors, the National Longitudinal Study of Adolescent Health (Add Health). The Add Health is a school-based, longitudinal study of the health-related behaviors of adolescents and their outcomes in young adulthood. In short, the study contains an in-school questionnaire administered to a nationally representative sample of students in grades 7 through 12 in 1994-95 and three in-home surveys that focus on a subsample of students in 1995 (Wave 1), and approximately one year (Wave 2) and then six years later (Wave 3). The fourth wave of the survey should be available for analysis later this year. The study began by using a clustered sampling design to ensure that the 80 high schools and 52 middle schools selected were representative of US schools with respect to region of country, urbanicity, size, type, and ethnicity. Eligible high schools included an 11th grade and enrolled more than 30 students. More than 70 percent of the originally sampled high schools participated. Each school that declined to participate was replaced by a school within the stratum.

For this paper, we focus on the In-School data collection, which utilized a self-administered instrument to more than 90,000 students in grades 7 through 12 in a 45- to 60-minute class period between September 1994 and April 1995. The questionnaire focused on topics including socio-demographic characteristics, family background, health status, risk behaviors, and friendship nominations. In particular, each student respondent was asked to identify up to 10 friends (5 males, 5 females) from the school's roster. Based on these nominations, social networks within each school can be constructed and characterized, linking the health behaviors of socially connected individuals.

Of the nearly 90,000 students in the schools originally surveyed, several reductions in the sample size were made in order to construct the analysis sample. First, nearly 4,500 students did not have individual identification numbers assigned. Nearly

12,000 students did not nominate any friends and 5,000 individuals nominated friends who were not able to be linked with other respondents due to nominations based on incomplete information ("nicknames" rather than names, or the nominated friend did not appear on the Add Heath school roster, etc.) These issues reduced the sample to approximately 66,000 respondents. In Appendix 1A, we examine the correlates of the individual dropped from the sample for these reasons. Briefly, race, gender, family structure, and missingness on other variables predicts sample selection to some extent, however health behaviors are not robust important predictors.

In this paper, our main focus is on individuals with same-sex/same-grade level friends, which reduces the sample to approximately 58,000 students.[3] One reason to focus on same-sex friends is that romantic relationships may be nominated as "friends". In addition, most previous studies of friendship networks also limit the network definition to same-sex friends. We limit our analysis to same-grade friends in order to use cross-cohort (grade) variation in friendship opportunities and choices, as we describe below. While our main focus is on same-sex friendship networks, we also present some evidence of opposite sex friendship networks to examine potential heterogeneity of effects and extend the literature in this direction. In order to retain sample size, we impute missing covariates, such as maternal education, and control for missingness, but we do not impute missing outcomes.

Table 1 presents descriptive statistics of the analysis sample and shows that approximately 34% of the sample reports smoking and 54% of the sample reports drinking alcohol. The average adolescent nominates 2.4 same-sex friends. In Appendix Table 2A, we stratify the descriptive statistics by race/ethnicity. As noted by other research, there is considerable variation in the composition of friendship networks; 84% of friend nominations by white students are to white students, while only 9% of friend nominations by black students are to white students, for example. We also show that the range of friendship network health behaviors includes individuals who have no smoking/drinking friends through individuals who have all smoking/drinking friends. In Table 2 we present the distribution of friends' health behaviors in the data.

[3] Of the 66,000 students, 4,300 do not nominate any same grade friends and 4,100 do not nominate any same-grade/same-gender friends (that is, they nominate same grade friends but no same-grade/same gender friends).

Identification Strategy

In this paper, we seek to estimate the causal effects of friends' health behaviors by overcoming the many empirical obstacles we outline above, including selection into networks, unobserved determinants of behaviors, and the joint determination of outcomes within a network. The intuition behind our approach is that we seek to form comparison groups based on information in the data that describes the friendship *options* of students as well as the students' *choices* of friends (given these options) following the premise that individuals who make similar decisions or have similar outcomes when facing the same set of options likely are very similar on both observable and unobservable attributes.

More specifically, we propose to begin our analysis by considering the following empirical model:

$$health_{ist} = \delta health_{jst} + X_{jst}\beta + X_{it}\beta + \varepsilon_{is} + \mu_{ist} \quad (1)$$

where $health$ indicates a particular health behavior, such as tobacco consumption, of individual *i* in social network *s* at time *t*. Health behaviors may be a function of the health behaviors of other individuals in the social network ($health_{jst}$), the individual's and group members' observable characteristics X, the individual's unobservable characteristics ε and an idiosyncratic error term μ.

As Manski (1993) demonstrates, even without correlations in social networks that are caused by unobservables, e.g. ε_{is} orthogonal to X_{js}, this model is intrinsically unidentified. By this we mean that there is insufficient information in the regression to estimate uniquely the parameters of interest (δ in particular). This occurs because group member characteristics that might explain the health of group members j and so act as instruments for health cannot be excluded from the second stage regression for the health behaviors of i because these attributes may just as reasonably directly influence i's behaviors (the reflection problem).[4] [5]

[4] For example, if one observes clustering of criminal behavior among friends whose parents have less education, even after controlling for all possible individual and environmental factors that might explain such clustering available in the data, we still cannot conclusively determine whether the clustering is caused because having friends whose parents have less education contributes to criminal behavior or individuals whose parents have less education are more likely to engage in criminal behavior and such criminal behavior influences the behavior of the individual's friends. See Brock and Durlauf (2001, 2006) for recent methodological progress on this problem.

Our strategy sorts students into clusters *c* based on comparing similar students who faced similar friendship options and made similar friendship choices. This sorting is based on observable (to the researcher) and unobservable characteristics. Therefore, as argued by Dale and Krueger, the inclusion of fixed effects for such clusters should assure that we are comparing students who are similar on both observables and unobservables, which, critically, breaks the correlation between peers' behaviors and a student's unobservable characteristics. Further, since all students in a cluster should have similar observable characteristics, the inclusion of the fixed effect also captures the observables associated with the students' peers while allowing effect of behavioral differences within cluster to identify the effect on friend behavior on individual behavior. This feature of the approach solves the empirical problem outlined above and isolates the causal effect of student behaviors on the behavior of their friends.

Specifically, define a cluster of individuals c_i in the same school who are observationally equivalent on observables *X* and choose observationally equivalent friends based on observables *X*. This structure implies that the individual and friendship group observables are the same within a cluster so that

$$X_{jst}\beta_1 + X_{it}\beta_2 = X_{lst}\beta_1 + X_{kt}\beta_2 \quad (2)$$

where *i* and *k* belong to the same cluster *c* and *l* are friends of individual *k*. Further, we assume that observationally equivalent individuals who face the same friendship options and make the same choices have the same value on the unobservable that influences health behavior. In other words, we assume that idiosyncratic factors driving differences in friendship choices within clusters are not correlated with behavior so that

$$\rho_{ic} = X_{jst}\beta_1 + X_{it}\beta_2 + \varepsilon_{is} = X_{lst}\beta_1 + X_{kt}\beta_2 + \varepsilon_{ks} = \rho_{kc} \quad (3)$$

The resulting specification is

$$health_{ist} = \delta health_{jst} + \rho_{ic} + \mu_{ist} \quad (4)$$

Empirical Specification

[5] As noted by Sacerdote (2001) and Bayer and Ross (2008), when social network effects are determined in part by unobservable characteristics, even random assignment cannot solve this identification problem. While random assignment breaks the correlation between the health behavior of j and i's unobservable characteristics, the coefficient estimate on the behavior of j is a composite of both the direct effect of peer's behaviors and the effect of peers' unobservable characteristics.

Our friendship clusters are based on students in the same school choosing sets of friends with very similar demographic attributes. As there is evidence that adolescents have strong preferences to befriend classmates based on age, gender, and race (Mayer and Puller 2008; Weinberg 2008), we will begin our analysis by flexibly controlling for these decisions in terms of friendship composition. In particular, in order to implement our approach, we begin by limiting friends to same-sex/same-grade nominations. In Appendix Table 3A, we present descriptive regression results that show the associations between nominating no same-sex/same grade friends for individuals that send nominations. The likelihood increases by grade and is smaller for more advantaged students. We find that the proportion of smokers in the grade (potential friends) is not related to these nomination patterns, however, individuals with drinking grademates are slightly more likely to nominate same-grade/same-gender friends (a 10 point increase in grademates drinking is associated with a 1 percentage point increase in the probability).

Next we must create our "individual type"-"friendship type clusters." Given a limited sample, there is clearly a trade-off between how restrictive we make our definitions of observationally similar individuals and of same friendship types. We begin by placing the most weight on obtaining very specific "friendship-type" clusters. The reason behind this focus in that most of our demographic variables are binary and so after controlling for individual-type on those variables very little information is left that can be used in our specification tests in order to examine whether peer attributes can explain predetermined student attributes. For example, we examine whether peer attributes can explain student race or ethnicity in a model that only controls for within school friendship types. However, in our final model specification, we include a whether the student is white and whether their mother is a college graduate in the creation of the friendship type clusters, and then for years of maternal education we can test whether peer within cluster variation can explain a student's own maternal education. In future versions of the paper, we will explore a broader set of individual controls for constructing individual-friendship type clusters.

The friendship clusters are based on the following exogenous characteristics of chosen friends, including (1) race (black vs. Hispanic vs. white vs. Asian vs. other) (2) maternal education (no college vs. some college vs. college graduate) (3) family structure

(living with mother vs. not living with mother) and (4) nativity (native vs. foreign born). Specifically, the number of friends chosen from each characteristic is used in the cluster. Importantly, our clusters are quite flexibly created, such that an individual who chooses five black friends is in a different cluster than an individual who chooses four black friends.[6] In yet another refinement of our cluster approach, in some analyses we also include grade levels-pairs within the clusters, so that 7th and 8th graders are compared to each other (and 9th/10th and 11th/12th) in order to move closer to the thought experiment outlined earlier.

Assuming a rich structure of clusters, as outlined above, will create single clusters of students—those students who have "unusual" friendship preferences. These single clusters will, implicitly, not contribute to the identification of the network effects estimates, as there will be no within-cluster variation to exploit. In Appendix Table 4A, we examine the predictors of being placed in a single cluster. In columns 1-4, we examine the basic clusters and columns 5-8 examine the within-school-grade-pair clusters. Asian students are 12 percentage points more likely to be placed in a single cluster, likely because they are more likely to nominate a large number of same-race friends; this limits the availability of other students with similar preferences in the data. Interestingly maternal education and the maternal caring index is not related to placement in a single cluster. We also find very small relationships between smoking and drinking status and placement in a single cluster.

A model that controls for school by friendship type by individual type fixed effects will still include variation in peer health behavior between students in the same grade and school. This variation raises the concern that the students who chose a friend with poor health behaviors differs on unobservables from students who selected friends with exactly the same attributes in the same grade, but among their options chose the students with better health behaviors. As discussed above, we can address this concern by restricting our comparisons to students in different grades who are observationally

[6] As an example, friendship cluster 15 could be created based on nominating four friends such that: friend A is white, has a college educated mother, lives with his mother, and is native born; friend B is white, has a mother with some college, lives with his mother ,and is native born; friend C is white, has a college educated mother, lives with his mother, and is foreign born; friend D is black, has a college educated mother, lives with mother, and is native born. Cluster 16 could be identical except the individual nominated four white friends instead of three white friends and one black friend; Cluster 17 could be identical to cluster 15 except all the nominated friends are native born.

equivalent on *X* and chose the same friendship set on the *X*'s. These students are unable to form the same own-grade friendships and so one student could not intentionally select away friends in their comparison group's friendship set. In order to accomplish this, we will randomly choose only one student in each grade from each friendship type cluster so that the estimated effect of peer behavior cannot be identified off of within grade variation. [7]

Naturally, the approach of using friendship cluster fixed effects as a solution to many of the empirical issues in estimating social network effects requires stronger assumptions than random assignment or even traditional cohort based studies of peer effects, but this strategy provides a significant payoff by potentially providing estimates of the effect of peer behaviors on individual behaviors that are not contaminated by the influence of individual unobservables or by the direct influence of peer observable characteristics.

Before presenting the main empirical results, we first outline supporting evidence for our approach by exploring friendship *options* and friendship *choices* in the data. In particular, we first show that, within high schools, friendship options are quite constrained across cohorts. This is evidence consistent with our notion of quasi-randomization of individuals to potential friends occurring based on the age of the respondent, which places them into grade-levels within schools. We then show evidence of the endogeneity of friendship in the data and suggestive evidence of the extent our approach reduces this endogeneity.

---

[7] As discussed, an illustration of our combined methodology is that we can compare two students who attend the same high school and each selected five African American, male friends in their same grade. This indicates that these two students faced similar friendship choices and also selected similar friends, given these choices. The difference between these two individuals who seem to have very similar preferences for friends is that one individual is in the 9th grade (and thus selects 9th grade friends) and the second student is in 10th grade in the same school (and thus selects 10th grade friends). We therefore leverage the fact that age has determined whether each student is in 9th or 10th grade in this specific school, and we argue that this "quasi-experiment" allows us to use the 9th grader as a counterfactual to the 10th grader when examining whether health behaviors of friends ($health_{jst}$) impacts own-health behavior outcomes ($health_{ist}$). Thus, we use these two students as the counterfactual for what would have happened had they been in a different grade in the same school, and thus had a different set of friends. We argue that this comparison technique addresses two of the empirical difficulties with estimating causal social network effects: selection of network members (friends) and unobserved causal factors. We address these difficulties by comparison individuals in the same environment (same school) and who, but for their assignments to different grade levels, would have chosen the same friends (randomization based on age).

Evidence of Variation in Friendship Options

As we outline above, we are able to extend our empirical analysis by combining the friendship cluster design of "matching on options and choices" approach with the unique design of our data. That is, the data contains multiple cohorts within each surveyed high school, which allows an additional source of variation to be utilized in our analysis. For example, we can combine our basic approach with the use of cross-cohort, within-school variation and in doing so are able to compare students who face similar friendship options (are in the same school) and make similar friendship choices. This extension relies heavily on the assumption that individuals who attend the same school, but different grades, have essentially the same "types" of friendship options.

To what extent do students in the same school face similar friendship options? Using the Add Health data, we show below in Table 3 that controlling for school and grade effects can predict over 95% of the variation in racial composition of potential friends (classmates) in the data. Likewise, controlling for school and grade predicts 93% of the variation in peers' maternal education level and 96% of the variation in classmate nativity. These findings suggest that students in different grades but who attend the same school have very similar friendship options based on race and family background of peers.

However, fortunately for our strategy, there is substantially more variation across cohort, within schools in unhealthy behaviors. Using the same regression analysis, our data show that we only predict 77% of peer smoking rates, 76% of exercise rates, and 81% of peer drinking rates. Thus, these results suggest that there is substantial variation in exposure to health behaviors of potential friends (classmates) even within school, while at the same time the friendship options based on race, maternal education, and nativity is nearly identical for students across grades within the same school. We use these features of our data to make comparisons within schools—students who face similar environments and make similar friendship choices but have different unhealthy behavior outcomes.

Evidence of Friendship Selection

The estimation approach we outline has been effectively and convincingly applied in the context of selection into higher education (Dale and Krueger 2002), as discussed above. Further, we can partially test the validity of the approach by examining whether students seem to be sorting into specific friendship patterns within our friendship clusters. Specifically, we recognize that given limited data our student friendship clusters will not perfectly control for student attributes. Thus, we test whether a student's own observable attributes correlate with the attributes of their friends within student clusters. Following the logic of Altonji, Elder, and Tabor (2005), if individuals do not sort on observables into friendships within clusters, it is very unlikely that they have sorted based on unobservable characteristics. For example, if we find no evidence of additional correlation between an individual's own parental education and the parental education of their friends after conditioning on the average level of correlation for all students in this cluster, which might include broader educational categories, then it is unlikely that students are sorting based on unobservable characteristics like the parents' involvement with the students' education or the parents' educational and academic expectations since those unobservable characteristics are likely correlated with parental education. Similar diagnostic tests have been used elsewhere (Bayer, Ross and Topa 2008; Bifulco, Fletcher and Ross 2008).

In Table 4, we present evidence from these diagnostic tests. Each set of rows examines the correlation between a different "outcome" (individual-level characteristic) and friend's characteristics. Columns add controls from left to right. The first column and row shows the correlation between whether an individual is of Hispanic ethnicity (vs non-Hispanic) and the average of his or her friends' maternal education levels (-0.03). Column 2 controls for school fixed effects and reduces the coefficient by 1/3. Column 3 uses friendship cluster fixed effects rather than school fixed effects and reduces the coefficient from column 1 by over two-thirds. Column 4 includes school fixed effects as well as cluster fixed effects and shows that cluster fixed effects reduce the correlation considerably. Column 5 controls for school X cluster fixed effects and reduces the coefficient to $1/10^{th}$ the size of the baseline regression. Column 6 adds grade-pairs to the clusters so that $7^{th}/8^{th}$, $9^{th}/10^{th}$, and $11^{th}/12^{th}$ graders are compared. Finally, column 7

adds individual characteristics to the cluster, including white race and whether the student's mother graduated from college.

The second row of coefficients in Table 4 examines the correlation between white race (vs. non-white) and the average maternal education level of selected friends and shows similar results to Row 1. In Row 3, we examine the correlation between own-maternal education and the average maternal education of friends. Here, the correlation is quite high—0.33—in the baseline specifications. As we add school X cluster fixed effects in column 5, the coefficients is reduced by more than two-thirds, but is still statistically significant. Finally, we include individual characteristics in the clusters definitions, and the correlation between own and friends' maternal education falls to 0.014 and is not statistically significant.

In a second set of balancing tests (Table 4B), we examine the correlations between individual characteristics and friends' health behaviors in order to further assess our ability to control for observables and unobservables in our estimation strategy. In the first row, we show that maternal education is highly associated with friends' drinking behaviors. However, when we control for clustering, the coefficient is reduced by 90% and is no longer statistically significant. In row 2, we find similar evidence from the correlation between maternal education and friends' smoking behaviors. This is suggestive evidence that our cluster controls are reducing endogeneity bias associated with students choosing their friends. Therefore, even prior to fully incorporating our cohort approach into the friendship type cluster design, there is no evidence to suggest that peer health behaviors can explain predetermined student attributes.

In row 3, we examine the correlation between a variable completely outside of our cluster definition—maternal caring. Again, we find that individuals with caring mothers are more likely to have friends with caring mothers. However, as we add cluster fixed effects in the final column, this correlation is reduced over 50% and is only statistically significant at the 10% level. This result is a strong test of the adequacy of our clusters and suggests a potential for minor violations but one cannot rule out the explanation that maternal education of peers actually has a causal influence on a student's report of maternal caring. Of course, we will control for maternal caring in our results, so any

residual correlations in unobservables between the respondent and his friends will be net of these controls and the cluster fixed effects.

Results

Same-Sex Friends

Table 5 presents estimates for adolescent smoking where same-sex/same-grade friends are used to define the friendship network. In Column 1, the baseline results suggest that increasing the share of friends who smoke by 10 percentage points would increase own-smoking by nearly 3.9 percentage points. Following previous research on smoking (Fletcher in press), we find that black students smoke less than white students, as do students with more highly educated mothers. We also show a steep rise in smoking at the transition between middle school and high school and then a plateau during high school.

In Column 2, we follow some previous literature and control for high school fixed effects; however this only reduces the coefficient from 0.388 to 0.368 for friends' smoking. In Column 3 we do not use school fixed effects, but instead use friendship cluster fixed effects. As discussed above, we create cluster fixed effects based on several aspects of the respondent's friendship nomination patterns, including (a) number of nominations (b) race of nominated friends (white vs. black vs Hispanic vs. Asian vs. other race), maternal education of nominated friends (college graduate vs. non college graduate), whether friend is native born, and whether friend lives with his/her mother. We choose these characteristics in part because they are exogenous characteristics of potential friends and several have been shown to be important predictors of friendships in other research (Mayer and Puller 2008, Marmaros and Sacerdote 2006). Altogether, this choice of cluster variables creates nearly 3,000 friendship clusters for the analysis. In addition, since we focus on same-sex/same-grade friendships, we are also implicitly using gender in the friendship clusters since we control for them in the analysis. It is important to also note that respondents are exactly matched based on these friendship choices in the analysis. With the inclusion of cluster fixed effects, the coefficient estimate mirrors that of the school fixed effects results (column 1 vs. column 3) from 0.39 to 0.37.

When we control for both school fixed effects and cluster fixed effects in Column 4, we get an estimate that implies increasing the proportion of friends who smoke by 10 points will increase own-smoking by 3.5 percentage points. In column five we control for school X cluster fixed effects, which results in an estimate of 3.1 percentage points. Finally, we limit the comparisons to those in adjacent grades (7/8, 9/10, 11/12), and then control for both individual and friendship types in our cluster definitions. The final estimate implies a 2.91 percentage point increase from a 10 point increase in smoking peers. Overall, we see a 25% reduction in the baseline estimate with our inclusion of individual-friendship type fixed effects. However, these changes are very small relative to the declines in estimates across the same model specifications for our balancing tests where the declines are typically on the order of 75 to 90 percent and the magnitudes of the final estimates are well below a 1% reduction for a 10 percent increase in peer behavior. As discussed above, as we control for richer cluster definitions, the sample size used to identify the coefficients is reduced due to "singleton clusters". In Appendix Table 5A, we show that the change in composition is not the explanation for our results by estimating the baseline results in Table 5 using the non-singleton sample across columns.

Table 6 examines drinking behaviors. Baseline results in column 1 suggest that a 10 percentage point increase in friends' drinking is associated with a 3.3 percentage point increase in own-drinking. Like the results for smoking, school fixed effects (added in column 2) reduce this association by a modest amount. Using the same cluster definition as in smoking, the results using friendship-cluster fixed effects (but not school fixed effects) in column 3 the coefficient is reduced slightly, suggesting that increasing friends' drinking by 10 points will increase own drinking by 3.2 percentage points. In column 4 we control for both school and friendship cluster fixed effects. The results suggest that increasing friends' drinking by 10 points increases own drinking by 2.9 percentage points. In column 5, we control school X cluster fixed effects, which lowers the coefficient to 2.6 percentage points, and in column 6 we add grade-pair comparisons, which reduces the coefficient to 2.4 and is our preferred estimate. The reduction in the estimate effect across columns is over 26%. These results suggest important effects of

friends' unhealthy behaviors on own-choices and also show that previous research likely produced inflated estimates of these effects.

In Table 7, we report estimates for the one-student-per-cluster specifications we outline above. A concern with results in previous tables is that we could be comparing two students *in the same grade* who have chosen different friends, conditional on the identical set of friend options. This type of comparison may exacerbate selection effects because the identification is coming from selection based on unobservables rather than across-cohort comparisons. Thus, we narrow our comparisons in Table 7 by randomly selecting one student per school-grade pair-cluster for the analysis for those grades with multiple students sharing a cluster. Results suggest that the selection effects serve to slightly bias our earlier coefficients toward zero, with our preferred estimates suggesting that an increase in friends' health behaviors by 10 points will increase own-smoking by 3.25 percentage points and own-drinking by 2.6 percentage points.

In Table 8, we examine gender and racial differences in the effects of same-sex friends. Results for both smoking and drinking suggest that the baseline social network effects are 1/3 higher for females than males. Interestingly, the gender gap shrinks considerably once controls are added. This is suggestive evidence that rather than females being more susceptible to peer pressure/social network effects, there is higher selection into friendships for females than males based on health behaviors. For the racially stratified results, we find evidence of larger social network effects for whites—the differentials are largely unaffected after we include our cluster fixed effects.

Opposite Sex Friends

We next extend our analysis to focus on opposite-sex friends. The effects are likely a combination of the influence of opposite sex friends as well as romantic partners, but represent a contribution to the literature because most studies focus on same-sex friends. The results in Table 9 suggest smaller influences from opposite-sex friends—a 10 point increase in friends' smoking is associated with a 2.3 percentage point increase in the likelihood of own-smoking. After controlling for "friendship types and options", the effect is reducted by 13% to a 2 percentage point increase. In Table 10, we estimate the effect of increasing friends' drinking by 10 points is associated with an increase of 2.1

percentage points in own-drinking. The effect is reduced by over 20% when controls are included. In Table 11, we examine the effects by gender and race. We find no evidence of differential effects by gender. The results by race suggests larger friendship network effects for white students and little evidence of effects for black students after including controls.

Extensions: Non-linear Effects, Age Effects, and Duration Effects

In Table 12, we examine the evidence of non-linear social network effects, first with the addition of a squared peer behavior term and then we break the effect in two ways: (1) categories based on the proportion of friends who smoke or drink and (2) categories based on the number of friends who smoke or drink. In column 1, we find evidence that the peer effect for smoking is large and increasing, however adding cluster fixed effects in column 2 eliminates the statistical significance of the squared term. In column 4 we break the categories of peer influence into: 0% (omitted category), 1%-49%, 50%, 51%-99%, and 100%. The results for smoking suggest an effect that is approximately linear. In columns 5 and 6 we examine the number of friends who smoke and find evidence that each additional smoking friend increases the likelihood of smoking by 10 percentage points. In column 6, using cluster fixed effects, we find a slightly larger effect that diminishes as an additional smoking friend is added. For drinking behaviors, we find somewhat similar results. The squared peer term is statistically significant in the baseline results but not after controlling for clusters. However, in the categories of peer influence we find some suggestive evidence that the effects of peers drinking may only be operative when 50% or more of a student's friends drink. We also find evidence of non-linear effects in the number of friends who drink.

In Table 13, we examine possible age and duration effects in the coefficient of interest. Column 1 presents the preferred result from Tables 5 and 6. We show in column 2 that the main result is unchanged if we limit the sample to those in grades 9-12. Columns 3 and 4 compare the effects of friends' health behaviors for students who have completed their first year at the school versus students who have been at the school longer. For both behaviors, we find evidence of that longer exposure increases the effect of friends' health behaviors.

Falsification Test

In Table 14, we seek to examine the performance of our cluster research design. Here we look for a health outcome that has no reasonable social contagion element, so we examine self reported frequency of chest pains in the previous month. In the OLS results in column 1, we find evidence that a 10% increase in the proportion of friends who report chest pain increases on reports of chest pain by 0.5 percentage points. We see that school fixed effects, which rely on within-school randomization of friends (e.g. Trogden et al. 2008) do not adequately eliminate the fictional social effects. Interestingly we find that the results are still statistically significant and important even with school fixed effects and cluster fixed effects in column 4. In column 5, we see the coefficient is reduced by over 50% from the baseline results when we add school-cluster fixed effects and eliminates the statistical significance. Adding school-cluster-grade-pair fixed effects in column 6 further reduces the coefficient to near zero. Thus these results suggest our final columns in previous tables are the best evidence on social network effects for smoking and drinking behaviors of adolescents.

Empirical Extension

Although not included in this draft, we plan on extending the methods in this paper in several directions. One of which is the method by which we force comparison within school, across cohorts. Rather than removing school fixed effects via a general mean differencing, which then compares all student outcomes in a school based on an average baseline for the school, we will calculate unique means for differencing from student information in each grade where the mean is based on all students in a friendship pattern cluster that are not in that particular grade. Further, this differences process also addresses a bias that arises in fixed effects models with a small numbers of students in each cluster. As noted in previous research (Bayer, Ross, and Topa 2008), leaving an individual in their own cluster for mean differencing creates a positive correlation between the fixed effect and the individual's idiosyncratic error, but dropping the individual creates a negative correlation because the cluster mean is no longer a random sample. By differencing based on students in a cluster from other grades, the mean is

based on a random sample of students from those grades and yet is not correlated with the student's idiosyncratic error.

Conclusions

While researchers typically examine peer effects by defining the peer group broadly, this paper focuses attention on actual friends and implements a new research design to study the effects of friend's health behaviors on own health behaviors for adolescents. The main idea is to combine a cross-cohort, within school design with controls for friendship options through high school fixed effects and friendship choices through the use of "friendship type" fixed effects. We show that in the Add Health data used in this paper, there is evidence that our design is successful in narrowing down relevant comparison groups by controlling for the friendship choices and friendship options of adolescents. Our initial estimates also suggest that all results are robust to the restriction of sample to one student per cluster per cohort, which assures that the model is only identified based on comparisons of students across clusters. While our methodology is able to overcome a number of difficult issues in the estimation of social network effects in order to identify social interactions in health behaviors, it is not without limitations. One such limitation is that we have yet to identify the direction of the causation within a cluster of students—which students in the cluster were the likely initiator of certain health behaviors, and which students were influenced by that behavior? We also need to assess robustness of the results to different definitions of friendship clusters.

Overall, our results suggest that friendship network effects are important in determining adolescent tobacco and alcohol use but are over-estimated in specifications that do not fully take into account the endogeneity of friendship selection by 20-30%, and we also find evidence that gender differences in social network effects are explained by selection bias. We present new evidence of the effects of opposite sex friends on health behaviors and also find racial differences in friendship network effects.

Literature Cited

Altonji, J G., Elder, TE, CR Tabor. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools." *Journal of Political Economy*, 2005 113: 151-184.

Bayer. P and S. Ross. Identifying individual and group effects in the presence of sorting: A neighborhood effects application. *National Bureau of Economic Research* Working Paper #12211, 2008

Bayer, P., Ross, S. and G. Topa. Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 2008, 116 (6): 1150-1196

Ballester, C., Calvó-Armengol, A. and Y. Zenou. Who's who in networks. Wanted: the key player. E*conometrica,* 2006, 74, 1403-1417

Bearman, P., Moody J., and K. Stovel. Chains of affection: the structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 2004, 110 (1): 44-91

Bifulco, R., Fletcher, J. and S. Ross. The effect of classmate characteristics on individual outcomes: evidence from the add health. Yale University Working Paper, 2008

Bramoulle Y. and B. Rogers. Diversity and popularity in social networks. Social Science Research Network, Working Paper, 2009.

Brock, William A. and Stephen N. Durlauf. 2006. Identification of Binary Choice Models with Social Interactions. *Journal of Econometrics*, forthcoming.

Brock, William A. and Steven N. Durlauf. 2001. Discrete choice with social interactions. *Review of Economic Studies*, 68, 235-260.

Brown, B. B., Dolcini, M. M. & Leventhal, A. (1997) Transformations in peer relationships at adolescence: implications for health-related behavior. In: Schulenberg, J.,Maggs, J. L. & Hurrelmann, K., eds. *Health Risks and Developmental Transitions During Adolescence*, pp. 161–189. New York: Cambridge University Press.

Calvó-Armengol, A., Patacchini, E. and Y. Zenou. Peer effects and social networks in education. *Review of Economic Studies*, in press

Christakis N, Fowler J. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine,* 2007, 357: 370-9.

Christakis N, Fowler J. The collective dynamics of smoking in a large social network. *New England Journal of Medicine,* 2008, 358: 249-58.

Clark, Andrew and Youenn Loheac. (2007). "'It Wasn't Me, It Was Them!' Social Influence in Risky Behavior by Adolescents." *Journal of Health Economics*

Cohen-Cole, E. and Fletcher, J.M. Detecting implausible social network effects in acne, height, and headaches: Longitudinal analysis. *British Medical Journal,* 2008a, 337: a2533.

Cohen-Cole, E. and Fletcher, J.M. Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics* 2008b, 27 (5): 1382-1387.

Cutler, D. and Glaeser, E. Social interactions and smoking. *National Bureau of Economic Research* Working Paper #13477, 2007

Conley, TG., C Hansen, and PE Rossi. Plausibly exogenous. Social Science Research

Network Working Paper, 2008
Dale, S. and A. Krueger. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics*, 2002, 117(4):1491-1527
Fletcher, J.M. Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics*, in press.
Fletcher, Jason M. (2008). "Peer Influences on Alcohol Consumption." Yale University Working Paper
Gaviria, Alejandro and Steven Raphael. (2001) "School-Based Peer Effects and Juvenile Behavior." *Review of Economics and Statistics*, Vol 83 (2)
Hanushek, E.A., Kain, J.F., Markman, J.M., Rivkin, S.G., 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics* 18 (5), 527–544.
Jackson, M.O. Social and Economic Networks. Princeton, NJ: Princeton University Press, 2008.
Liu X, Lee LF, Kagel J. Dynamic discrete choice models with lagged social interactions: With an application to a signaling game experiment. Ohio State: Mimeo, 2006.
Lundborg, Peter. (2006). "Having the Wrong Friends? Peer Effects in Adolescent Substance Use." *Journal of Health Economics*, Vol 25: 214-233
Manski, Charles. (1993). "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*. 60
Manski, Charles. (1995). Identification Problems in the Social Sciences. Harvard University Press: Cambridge, MA
Manski, Charles. (2000). "Economic Analysis of Social Interactions." *Journal of Economic Perspectives*. 14:3
Mayer. A and S. Puller. The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics*, 2008, 92 (1-2): 329-347
Powell, Lisa, John Taurus, and Hana Ross. "The Importance of Peer Effects, Cigarette Prices, and Tobacco Control Policies for Youth Smoking Behavior." *Journal of Health Economics,* 2005, Vol 24, pp. 950-968.
Renna, Francesco, Irina Grafova, and Nidhi Thakur. (2008). "The Effects of Friends on Adolescent Body Weight." Economics & Human Biology Volume 6, Issue 3, Pages 377-387
Sacerdote. B. Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 2001, 116 92): 681-704
Weinberg, B. Social interactions and endogenous associations. Ohio State University Working Paper, 2008

Table 1
Descriptive Statistics
Add Health
Analysis Sample From In School Survey: Same Grade/Same Sex Friends

Variable	Obs	Mean	Std Dev	Min	Max
Smoke	62811	0.35	0.48	0	1
Drink	62674	0.54	0.50	0	1
Get Drunk	62307	0.30	0.46	0	1
Exercise	59991	2.28	1.20	0	4
Any Exercise	59991	0.95	0.22	0	1
Male	65495	0.47	0.50	0	1
White	65855	0.59	0.49	0	1
Hispanic	65855	0.14	0.35	0	1
Black	65855	0.18	0.38	0	1
Asian	65855	0.06	0.23	0	1
Live with Mom	64675	0.93	0.26	0	1
Maternal Years of Education	65855	13.41	2.33	0	18
Maternal Caring Scale	65855	4.78	0.61	1	5
Native Born	64164	0.92	0.28	0	1
Grade = 7	65456	0.14	0.35	0	1
Grade = 8	65456	0.14	0.35	0	1
Grade = 9	65456	0.21	0.41	0	1
Grade = 10	65456	0.19	0.40	0	1
Grade = 11	65456	0.17	0.37	0	1
Grade = 12	65456	0.15	0.36	0	1
Missing	65855	0.43	0.49	0	1
Number of Nominations	65855	2.41	1.53	0	5
Proportion White	57278	0.60	0.43	0	1
Proportion Black	57278	0.17	0.35	0	1
Proportion Hispanic	57278	0.13	0.29	0	1
Proportion Asian	57278	0.06	0.19	0	1
Proportion Other Race	57278	0.04	0.14	0	1
Proportion Mom Less High School	45427	0.15	0.28	0	1
Proportion Mom Some College	45427	0.18	0.28	0	1
Proportion of Mom College Grad	65855	0.35	0.31	0	1
Proportion Native	55509	0.92	0.22	0	1
Proportion Live with Mom	55794	0.93	0.18	0	1

Table 2
Distribution of Health Behaviors in Friendship Networks

% Smoke	Freq.	Percent	Cum.	% Drink	Freq.	Percent	Cum.
Same Sex Friends							
0.00	22,994	42.51	42.51	0.00	12,509	23.18	23.18
0.10				0.10			
0.20	1,534	2.84	45.34	0.20	931	1.73	24.91
0.30	7,270	13.44	58.78	0.30	5,542	10.27	35.18
0.40	1,154	2.13	60.91	0.40	1,064	1.97	37.15
0.50	7,146	13.21	74.12	0.50	7,713	14.3	51.45
0.60	770	1.42	75.55	0.60	1,135	2.1	53.55
0.70	2651	4.9	80.45	0.70	3,774	6.99	60.55
0.80	1,748	3.23	83.68	0.80	3,440	6.38	66.92
0.90				0.90			
1.00	8,830	16.32	100	1.00	17,847	33.08	100
Total	54,097	100		Total	53,955	100	
% Smoke	Freq.	Percent	Cum.	% Drink	Freq.	Percent	Cum.
Opposite Sex Friends							
0.00	15,965	43.48	43.48	0.00	8,516	23.26	23.26
0.10				0.10			
0.20	590	1.61	45.11	0.20	328	0.9	24.16
0.30	3,940	10.73	55.84	0.30	2,646	7.23	31.39
0.40	512	1.39	57.23	0.40	434	1.19	32.57
0.50	4,920	13.4	70.63	0.50	5,141	14.04	46.61
0.60	371	1.01	71.64	0.60	481	1.31	47.92
0.70	1653	4.5	76.15	0.70	2,360	6.44	54.37
0.80	859	2.34	78.49	0.80	1,643	4.49	58.86
0.90				0.90			
1.00	7,896	21.51	100	1.00	15,063	41.14	100
Total	36,706	100		Total	36,612	100	

Table 3
Variation in Friendship Options

Peer Variable	R-squared
% Maternal College Graduate	92.5%
% Black	97.2%
% Hispanic	97.4%
% White	
% Asian	93.8%
% Native Born	96.1%
Mean Maternal Caring Scale	55.1%
% Smoke Cigarettes	76.5%
% Drink Alcohol	80.9%
% Exercise	75.8%

Notes: The results reported indicate the R-squared from a regression of a complete set of school-level and grade-level dummy variables on the grade-level measure of peer characteristics or peer health behaviors N~65,000

Table 4
Balancing Tests of Friendship Sorting

Outcome	Hispanic	Hispanic	Hispanic	Hispanic	Hispanic	Hispanic	Hispanic
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster	School-GradePair-Cluster-X
Friends' Maternal Education	-0.032***	-0.022***	-0.003*	-0.004**	-0.003	-0.003	
	(0.006)	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	
Observations	65456	65456	65456	65456	65456	65456	
R-squared	0.027	0.306	0.425	0.488	0.696	0.725	
Outcome	White	White	White	White	White	White	White
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster	School-GradePair-Cluster-X
Friends' Maternal Education	0.026***	0.019***	0.001	0.003*	0.002	0.002	
	(0.007)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	
Observations	65495	65456	65456	65456	65456	65456	
R-squared	0.003	0.009	0.372	0.524	0.570	0.752	
Outcome	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster	School-GradePair-Cluster-X
Friends' Maternal Education	0.335***	0.202***	0.111***	0.067***	0.036*	0.036*	0.014
	(0.026)	(0.021)	(0.015)	(0.013)	(0.020)	(0.021)	(0.017)
Observations	65456	65456	65456	65456	65456	65456	65456
R-squared	0.061	0.123	0.202	0.243	0.530	0.586	0.869

Each set of rows and each column displays coefficients from separate regressions. All regressions control for grade-level fixed effects.

Table 4B

Balancing Tests of Friendship Sorting (Health Behaviors)

Outcome	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster	School-GradePair-Cluster-X
Friends' Drinking	-0.245***	-0.281***	-0.134***	-0.192***	-0.180***	-0.169***	-0.027
	(0.057)	(0.034)	(0.034)	(0.027)	(0.049)	(0.055)	(0.057)
Observations	53895	53895	53895	53895	53895	53895	53895
R-squared	0.002	0.112	0.229	0.271	0.603	0.665	0.915
Outcome	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education	Maternal Education
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster	School-GradePair-Cluster-X
Friends' Smoking	-0.304***	-0.334***	-0.208***	-0.244***	-0.239***	-0.236***	-0.071
	(0.073)	(0.048)	(0.037)	(0.039)	(0.059)	(0.067)	(0.061)
Observations	54027	54027	54027	54027	54027	54027	54027
R-squared	0.003	0.113	0.230	0.271	0.602	0.665	0.915
Outcome	Maternal Caring	Maternal Caring	Maternal Caring	Maternal Caring	Maternal Caring	Maternal Caring	Maternal Caring
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster	School-GradePair-Cluster-X
Friends' Maternal Caring	0.089***	0.079***	0.081***	0.071***	0.060***	0.053***	0.046*
	(0.008)	(0.008)	(0.009)	(0.009)	(0.016)	(0.019)	(0.025)
Observations	51017	51017	51017	51017	51017	51017	51017
R-squared	0.006	0.013	0.135	0.140	0.509	0.581	0.707

Each set of rows and each column displays coefficients from separate regressions. All regressions control for grade-level fixed effects.

Table 5
Friendship Network Effects on Smoking

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Friends	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster	School-GradePair-Cluster-X
% Smoke	0.388***	0.368***	0.372***	0.350***	0.310***	0.299***	0.291***
	(0.010)	(0.011)	(0.012)	(0.012)	(0.019)	(0.023)	(0.027)
Male	-0.006	-0.010**	-0.012**	-0.016***	-0.007	-0.007	-0.008
	(0.005)	(0.004)	(0.006)	(0.005)	(0.010)	(0.012)	(0.019)
Hispanic	-0.023**	0.004	0.000	0.009	-0.001	-0.006	-0.014
	(0.011)	(0.010)	(0.011)	(0.011)	(0.021)	(0.027)	(0.051)
Black	-0.088***	-0.102***	-0.092***	-0.097***	-0.092***	-0.097***	-0.140***
	(0.009)	(0.009)	(0.013)	(0.013)	(0.028)	(0.033)	(0.047)
Asian	-0.087***	-0.082***	-0.063***	-0.066***	-0.089***	-0.098***	-0.124**
	(0.009)	(0.011)	(0.011)	(0.012)	(0.025)	(0.027)	(0.049)
Maternal Education	-0.005***	-0.005***	-0.004***	-0.004***	-0.006***	-0.006***	-0.004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.006)
Maternal Caring Index	-0.072***	-0.073***	-0.071***	-0.071***	-0.070***	-0.074***	-0.072***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.008)	(0.009)	(0.014)
Native Born	0.060***	0.053***	0.045***	0.047***	0.055**	0.060***	0.056
	(0.013)	(0.009)	(0.014)	(0.012)	(0.023)	(0.022)	(0.036)
Missing	0.016***	0.021***	0.010*	0.013**	0.017	0.016	0.032
	(0.005)	(0.005)	(0.006)	(0.006)	(0.011)	(0.014)	(0.024)
Constant	0.487***	0.498***	0.480***	0.145***	0.508***	0.628***	0.601***
	(0.026)	(0.024)	(0.029)	(0.034)	(0.056)	(0.060)	(0.122)
Observations	50970	50970	50970	50970	50970	50970	50959
R-squared	0.136	0.143	0.244	0.252	0.579	0.649	0.761

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Additional Controls: Grade dummies

Table 6
Friendship Network Effects on Drinking

Outcome	Drink	Drink	Drink	Drink	Drink	Drink	Drink
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Friends	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster	School-GradePair-Cluster-X
% Drink	0.331***	0.303***	0.318***	0.288***	0.257***	0.244***	0.241***
	(0.011)	(0.012)	(0.013)	(0.013)	(0.019)	(0.022)	(0.029)
Male	0.009**	0.006	0.006	0.005	0.009	0.008	0.017
	(0.004)	(0.004)	(0.005)	(0.005)	(0.010)	(0.011)	(0.019)
Hispanic	0.021**	0.030***	0.038***	0.038***	0.049**	0.049*	0.075*
	(0.010)	(0.009)	(0.011)	(0.012)	(0.023)	(0.027)	(0.045)
Black	-0.035***	-0.047***	-0.066***	-0.070***	-0.070**	-0.068**	-0.094**
	(0.011)	(0.010)	(0.015)	(0.015)	(0.027)	(0.029)	(0.046)
Asian	-0.107***	-0.118***	-0.083***	-0.093***	-0.084***	-0.085***	-0.085
	(0.011)	(0.012)	(0.013)	(0.013)	(0.026)	(0.032)	(0.057)
Maternal Education	-0.004***	-0.006***	-0.004***	-0.006***	-0.006***	-0.007***	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.007)
Maternal Caring Index	-0.065***	-0.065***	-0.065***	-0.065***	-0.063***	-0.063***	-0.067***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)	(0.009)	(0.013)
Native Born	0.079***	0.078***	0.069***	0.073***	0.076***	0.081***	0.103***
	(0.013)	(0.012)	(0.012)	(0.012)	(0.020)	(0.026)	(0.036)
Missing	0.004	0.008	0.006	0.008	0.007	0.010	0.009
	(0.005)	(0.005)	(0.006)	(0.006)	(0.012)	(0.014)	(0.028)
Constant	0.482***	0.518***	0.490***	0.107***	0.533***	0.736***	0.649***
	(0.027)	(0.027)	(0.029)	(0.038)	(0.053)	(0.060)	(0.123)
Observations	50730	50730	50730	50730	50730	50730	50728
R-squared	0.149	0.159	0.267	0.277	0.606	0.671	0.780

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Additional Controls: Grade dummies

Table 7
Friendship Network Effects on Smoking and Drinking
Results for One-Student-Per-Cluster Comparisons

Outcome	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS
Fixed Effects	None	School-Cluster-GradePair-X	School-Cluster-GradePair-X
Sample	Full	Full	One Per Cluster
% Smoke	0.389***	0.291***	0.325***
	(0.010)	(0.027)	(0.031)
Obs	50951	50951	40449
R-squared	0.136	0.761	0.780

Outcome	Drink	Drink	Drink
Specification	OLS	OLS	OLS
Fixed Effects	None	School-Cluster-GradePair-X	School-Cluster-GradePair-X
Sample	Full	Full	One Per Cluster
% Drink	0.333***	0.241***	0.262***
	(0.011)	(0.029)	(0.034)
Obs	50711	50711	40247
R-squared	0.150	0.779	0.795

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Same controls as Table 6.

Table 8
Racial and Gender Differences for Same-Sex Friendship Networks

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster
Females						
% Smoke	0.430***	0.405***	0.407***	0.381***	0.314***	0.291***
	(0.013)	(0.014)	(0.016)	(0.017)	(0.031)	(0.037)
Males						
% Smoke	0.334***	0.308***	0.315***	0.290***	0.270***	0.260***
	(0.013)	(0.014)	(0.015)	(0.015)	(0.029)	(0.036)
White						
% Smoke	0.436***	0.416***	0.421***	0.397***	0.357***	0.347***
	(0.010)	(0.011)	(0.011)	(0.011)	(0.019)	(0.023)
Black						
% Smoke	0.201***	0.159***	0.175***	0.136***	0.080*	0.074
	(0.022)	(0.023)	(0.022)	(0.023)	(0.047)	(0.062)
Hispanic						
% Smoke	0.317***	0.283***	0.276***	0.252***	0.240***	0.230***
	(0.015)	(0.013)	(0.016)	(0.015)	(0.039)	(0.055)

Outcome	Drink	Drink	Drink	Drink	Drink	Drink
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-GradePair-Cluster
Females						
% Drink	0.363***	0.333***	0.347***	0.315***	0.268***	0.237***
	(0.013)	(0.014)	(0.015)	(0.016)	(0.027)	(0.033)
Males						
% Drink	0.290***	0.257***	0.277***	0.244***	0.218***	0.212***
	(0.015)	(0.015)	(0.016)	(0.016)	(0.028)	(0.032)
White						
% Drink	0.385***	0.353***	0.373***	0.337***	0.311***	0.294***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.019)	(0.023)
Black						
% Drink	0.201***	0.167***	0.188***	0.154***	0.142***	0.133***
	(0.017)	(0.019)	(0.018)	(0.020)	(0.037)	(0.046)
Hispanic						
% Drink	0.240***	0.214***	0.189***	0.166***	0.136**	0.118
	(0.021)	(0.021)	(0.030)	(0.030)	(0.055)	(0.074)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9
Friendship Networks Effects of Smoking: Opposite Sex Friends

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Friends	Opp Sex/ Same Grade	Opp Sex/ Same Grade	Opp Sex/ Same Grade	Opp Sex/ Same Grade	Opp Sex/ Same Grade	Opp Sex/ Same Grade
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-Grade Pair-Cluster
% Smoke	0.231***	0.205***	0.229***	0.203***	0.207***	0.201***
	(0.010)	(0.010)	(0.011)	(0.011)	(0.019)	(0.022)
Male	-0.009	-0.010	-0.018*	-0.018*	-0.012	-0.008
	(0.009)	(0.009)	(0.010)	(0.010)	(0.017)	(0.020)
Hispanic	-0.041***	-0.003	-0.023	-0.006	-0.016	-0.022
	(0.016)	(0.012)	(0.014)	(0.013)	(0.023)	(0.028)
Black	-0.105***	-0.120***	-0.091***	-0.105***	-0.096***	-0.109***
	(0.011)	(0.011)	(0.014)	(0.015)	(0.031)	(0.036)
Asian	-0.097***	-0.093***	-0.054***	-0.057***	-0.065*	-0.072
	(0.013)	(0.016)	(0.017)	(0.018)	(0.036)	(0.045)
Maternal Education	-0.007***	-0.008***	-0.006***	-0.007***	-0.009***	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Maternal Caring	-0.079***	-0.079***	-0.078***	-0.078***	-0.077***	-0.074***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.010)	(0.011)
Native Born	0.071***	0.059***	0.057***	0.057***	0.051*	0.049
	(0.018)	(0.013)	(0.020)	(0.017)	(0.031)	(0.036)
Grade = 8	0.068***	0.071***	0.070***	0.073***	0.047*	0.061**
	(0.009)	(0.009)	(0.010)	(0.011)	(0.026)	(0.031)
Grade = 9	0.125***	0.137***	0.133***	0.148***	0.151***	0.000
	(0.011)	(0.018)	(0.012)	(0.021)	(0.039)	(0.000)
Grade = 10	0.126***	0.144***	0.127***	0.146***	0.161***	0.015
	(0.012)	(0.018)	(0.014)	(0.021)	(0.040)	(0.022)
Grade = 11	0.129***	0.147***	0.134***	0.153***	0.154***	0.000
	(0.011)	(0.019)	(0.013)	(0.022)	(0.039)	(0.000)
Grade = 12	0.143***	0.160***	0.147***	0.165***	0.174***	0.008
	(0.012)	(0.019)	(0.014)	(0.022)	(0.039)	(0.025)
Missing	0.039***	0.045***	0.016**	0.019**	0.025	0.026
	(0.006)	(0.006)	(0.008)	(0.008)	(0.015)	(0.020)
Constant	0.580***	0.597***	0.570***	0.243***	0.600***	0.707***
	(0.033)	(0.034)	(0.035)	(0.064)	(0.071)	(0.079)
Observations	34487	34487	34487	34487	34487	34487
R-squared	0.079	0.093	0.221	0.234	0.592	0.660

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 10
Friendship Networks Effects of Drinking: Opposite Sex Friends

Outcome	Drink	Drink	Drink	Drink	Drink	Drink
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Friends	Opp Sex/ Same Grade	Opp Sex/ Same Grade	Opp Sex/ Same Grade	Opp Sex/ Same Grade	Opp Sex/ Same Grade	Opp Sex/ Same Grade
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-Grade Pair-Cluster
% Smoke	0.212***	0.181***	0.215***	0.182***	0.182***	0.169***
	(0.011)	(0.011)	(0.013)	(0.012)	(0.020)	(0.024)
Male	0.023***	0.022***	0.018**	0.019**	0.016	0.016
	(0.007)	(0.007)	(0.008)	(0.007)	(0.013)	(0.015)
Hispanic	0.017	0.022**	0.022	0.021*	0.040	0.036
	(0.013)	(0.010)	(0.014)	(0.012)	(0.024)	(0.033)
Black	-0.047***	-0.065***	-0.058***	-0.072***	-0.053*	-0.061*
	(0.013)	(0.012)	(0.015)	(0.016)	(0.029)	(0.033)
Asian	-0.116***	-0.136***	-0.074***	-0.093***	-0.081**	-0.091**
	(0.014)	(0.013)	(0.016)	(0.017)	(0.036)	(0.044)
Maternal Education	-0.007***	-0.010***	-0.006***	-0.009***	-0.009***	-0.009***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)
Maternal Caring	-0.065***	-0.065***	-0.064***	-0.064***	-0.061***	-0.058***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.009)	(0.012)
Native Born	0.110***	0.107***	0.107***	0.109***	0.121***	0.132***
	(0.018)	(0.016)	(0.018)	(0.017)	(0.026)	(0.032)
Grade = 8	0.097***	0.102***	0.096***	0.102***	0.061**	0.062**
	(0.008)	(0.009)	(0.011)	(0.012)	(0.026)	(0.031)
Grade = 9	0.192***	0.201***	0.198***	0.204***	0.225***	-0.052**
	(0.013)	(0.014)	(0.015)	(0.016)	(0.030)	(0.022)
Grade = 10	0.244***	0.258***	0.244***	0.255***	0.276***	0.000
	(0.014)	(0.015)	(0.017)	(0.017)	(0.032)	(0.000)
Grade = 11	0.286***	0.301***	0.286***	0.298***	0.297***	0.000
	(0.014)	(0.016)	(0.016)	(0.017)	(0.031)	(0.000)
Grade = 12	0.326***	0.343***	0.327***	0.341***	0.357***	0.053**
	(0.015)	(0.015)	(0.018)	(0.018)	(0.032)	(0.021)
Missing	0.014**	0.018***	0.004	0.006	0.005	0.007
	(0.005)	(0.005)	(0.008)	(0.007)	(0.018)	(0.022)
Constant	0.523***	0.577***	0.517***	0.039	0.523***	0.729***
	(0.034)	(0.031)	(0.036)	(0.059)	(0.068)	(0.073)
Observations	34334	34334	34334	34334	34334	34334
R-squared	0.111	0.128	0.260	0.276	0.626	0.687

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 11
Racial and Gender Differences for Opposite-Sex Friendship Networks

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-Grade Pair-Cluster
Females						
% Smoke	0.234***	0.215***	0.233***	0.214***	0.208***	0.218***
	(0.011)	(0.011)	(0.012)	(0.013)	(0.025)	(0.029)
Males						
% Smoke	0.229***	0.216***	0.227***	0.211***	0.228***	0.227***
	(0.012)	(0.011)	(0.013)	(0.013)	(0.028)	(0.035)
White						
% Smoke	0.263***	0.238***	0.260***	0.233***	0.239***	0.231***
	(0.010)	(0.010)	(0.011)	(0.011)	(0.023)	(0.028)
Black						
% Smoke	0.115***	0.072***	0.124***	0.079***	0.061	0.057
	(0.016)	(0.017)	(0.021)	(0.022)	(0.056)	(0.065)
Hispanic						
% Smoke	0.180***	0.148***	0.166***	0.139***	0.159***	0.183**
	(0.020)	(0.019)	(0.027)	(0.030)	(0.061)	(0.075)
Outcome	Drink	Drink	Drink	Drink	Drink	Drink
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	None	School	Cluster	School,Cluster	School-Cluster	School-Grade Pair-Cluster
Females						
% Drink	0.209***	0.187***	0.212***	0.188***	0.187***	0.170***
	(0.012)	(0.012)	(0.014)	(0.013)	(0.032)	(0.037)
Males						
% Drink	0.217***	0.188***	0.223***	0.193***	0.191***	0.181***
	(0.015)	(0.014)	(0.016)	(0.015)	(0.031)	(0.034)
White						
% Drink	0.257***	0.218***	0.259***	0.218***	0.228***	0.216***
	(0.013)	(0.012)	(0.013)	(0.012)	(0.021)	(0.025)
Black						
% Drink	0.104***	0.070***	0.104***	0.073***	0.048	0.045
	(0.020)	(0.021)	(0.027)	(0.028)	(0.049)	(0.053)
Hispanic						
% Drink	0.160***	0.146***	0.149***	0.133***	0.151***	0.115**
	(0.023)	(0.023)	(0.033)	(0.035)	(0.050)	(0.048)

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 12
Non-Linear Effects of Friends' Health Behaviors: Smoking

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Friends	Same Sex/ Same Grade	Same Sex /Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade
Fixed Effects	None	School-GradePair-Cluster	None	School-GradePair-Cluster	None	School-GradePair-Cluster
% Behavior	0.314***	0.267***				
	(0.021)	(0.092)				
% Behavior Squared	0.081***	0.024				
	(0.021)	(0.097)				
0% Behavior			Omitted	Omitted		
1%-49% Behavior			0.057***	0.049		
			(0.006)	(0.035)		
50% Behavior			0.186***	0.139***		
			(0.008)	(0.028)		
51%-99% Behavior			0.292***	0.245***		
			(0.010)	(0.037)		
100% Behavior			0.376***	0.285***		
			(0.010)	(0.027)		
0 Friends					Omitted	Omitted
1 Friend					0.150***	0.175***
					(0.006)	(0.018)
2 Friend					0.258***	0.334***
					(0.008)	(0.031)
3 Friend					0.352***	0.454***
					(0.013)	(0.052)
4 Friend					0.464***	0.540***
					(0.017)	(0.099)
5 Friend					0.531***	0.674***
					(0.037)	(0.231)
Constant	0.488***	0.639***	0.493***	0.643***	0.521***	0.578***
	(0.026)	(0.128)	(0.026)	(0.130)	(0.027)	(0.130)
Observations	50951	50951	50951	50951	50951	50951
R-squared	0.137	0.761	0.138	0.761	0.115	0.762

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Additional controls same as previous tables.

Table 12B
Non-Linear Effects of Friends' Health Behaviors: Drinking

Outcome	Drink	Drink	Drink	Drink	Drink	Drink
Specification	OLS	OLS	OLS	OLS	OLS	OLS
Friends	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade	Same Sex/ Same Grade
Fixed Effects	None	School-GradePair-Cluster	None	School-GradePair-Cluster	None	School-GradePair-Cluster
% Behavior	0.243***	0.164				
	(0.026)	(0.108)				
% Behavior Squared	0.087***	0.076				
	(0.024)	(0.105)				
0% Behavior			Omitted	Omitted		
1%-49% Behavior			0.031***	-0.012		
			(0.008)	(0.037)		
50% Behavior			0.144***	0.113***		
			(0.010)	(0.033)		
51%-99% Behavior			0.228***	0.161***		
			(0.010)	(0.041)		
100% Behavior			0.314***	0.233***		
			(0.011)	(0.028)		
0 Friends					Omitted	Omitted
1 Friend					0.144***	0.161***
					(0.008)	(0.023)
2 Friend					0.204***	0.267***
					(0.010)	(0.037)
3 Friend					0.279***	0.385***
					(0.010)	(0.050)
4 Friend					0.351***	0.529***
					(0.015)	(0.073)
5 Friend					0.405***	0.529***
					(0.017)	(0.132)
Constant	0.488***	0.743***	0.498***	0.750***	0.514***	0.549***
	(0.026)	(0.060)	(0.026)	(0.060)	(0.029)	(0.125)
Observations	50728	50728	50728	50728	50711	50711
R-squared	0.149	0.671	0.150	0.672	0.133	0.779

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Additional controls same as previous tables.

Table 13
Heterogeneity of Effects: Duration and Age

Outcome	Smoke	Smoke	Smoke	Smoke
Specification	OLS	OLS	OLS	OLS
Fixed Effects	School-Cluster-GradePair-X	School-Cluster-GradePair-X	School-Cluster-GradePair-X	School-Cluster-GradePair-X
Sample	Full	High School	1st Year Student	1+ Year Student
% Smoke	0.291***	0.295***	0.232***	0.295***
	(0.027)	(0.028)	(0.058)	(0.035)
Obs	50951	37026	13634	37317
R-squared	0.761	0.734	0.854	0.780

Outcome	Drink	Drink	Drink	Drink
Specification	OLS	OLS	OLS	OLS
Fixed Effects	School-Cluster-GradePair-X	School-Cluster-GradePair-X	School-Cluster-GradePair-X	School-Cluster-GradePair-X
Females	Full	High School	1st Year Student	1+ Year Student
% Drink	0.241***	0.240***	0.206***	0.238***
	(0.029)	(0.032)	(0.070)	(0.033)
Obs	50711	36831	13545	37166
R-squared	0.779	0.740	0.859	0.799

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Additional controls same as previous tables.

Table 14
Falsification Tests of Cluster Performance
The Case of Chest Pains

Outcome	Chest Pain	Chest Pain	Chest Pain	Chest Pain	Chest Pain	Chest Pain	Chest Pain
Group	None	School	Cluster	School/ Cluster	School-Cluster	School-Cluster-Cohort	School-Cluster-Cohort-Xs
% Chest Pain	0.056***	0.047***	0.048***	0.040***	0.024	0.015	0.015
	(0.009)	(0.009)	(0.010)	(0.010)	(0.016)	(0.019)	(0.030)
Observations	49855	49855	49855	49855	49855	49855	49855
R-squared	0.021	0.025	0.148	0.152	0.534	0.609	0.742

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Tables

Table 1A

Predictors of Dropped Sample

Outcome	Means	No ID	No Friend Nominations	No Found Nominations	Any Drop
Means		0.047	0.14	0.2	0.24
Age	14.99	-0.001	0.022***	0.007	0.024***
		(0.003)	(0.005)	(0.005)	(0.009)
Male	0.502	0.006***	0.068***	0.016***	0.078***
		(0.002)	(0.006)	(0.003)	(0.008)
Hispanic	0.155	0.000	0.038***	0.007	0.038
		(0.012)	(0.010)	(0.017)	(0.027)
Black	0.19	-0.009	0.053***	0.017	0.053*
		(0.012)	(0.013)	(0.019)	(0.029)
Asian	0.056	-0.009	0.021**	-0.002	0.010
		(0.015)	(0.009)	(0.016)	(0.026)
Native Born	0.9	-0.003	-0.022***	-0.026***	-0.041***
		(0.005)	(0.006)	(0.008)	(0.011)
Live with Mom	0.92	0.025***	0.108***	0.052***	0.145***
		(0.008)	(0.015)	(0.014)	(0.018)
Mom Education	13.36	0.000	-0.001	-0.000	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Mom Native	0.82	0.001	-0.006	-0.014**	-0.017**
		(0.005)	(0.007)	(0.005)	(0.008)
Smoke	0.36	0.007	0.006*	0.015**	0.023**
		(0.006)	(0.003)	(0.007)	(0.010)
Drink	0.55	0.002	-0.016***	0.004	-0.010
		(0.003)	(0.003)	(0.004)	(0.007)
Age Missing	0.004	0.035*	0.123***	-0.006	0.118***
		(0.019)	(0.031)	(0.029)	(0.033)
Male Missing	0.008	0.054***	0.067***	0.007	0.098***
		(0.016)	(0.023)	(0.015)	(0.022)
Native Born Missing	0.03	0.011**	0.065***	0.013*	0.071***
		(0.005)	(0.009)	(0.007)	(0.010)
Live with Mom Missing	0.03	0.004	0.116***	-0.029*	0.071***
		(0.008)	(0.015)	(0.015)	(0.016)
Mom Education Missing	0.22	0.002	0.049***	0.011***	0.053***
		(0.003)	(0.005)	(0.003)	(0.006)
Mom Native Missing	0.13	0.032***	0.117***	0.060***	0.164***
		(0.007)	(0.014)	(0.011)	(0.015)
Smoke Missing	0.07	0.007	0.114***	-0.009	0.092***
		(0.009)	(0.017)	(0.011)	(0.018)
Drink Missing	0.08	0.023**	0.065***	0.005	0.074***
		(0.009)	(0.018)	(0.012)	(0.019)
Observations		89047	84834	73393	89047
R-squared		0.007	0.081	0.011	0.061

Notes: Grade fixed effects controls (not significant). "No ID" is a binary variable indicating whether the respondent received an identification number in the survey. "No Friend Nominations" is a binary variable indicating whether the respondent made zero friend nominations. "No Found Nominations" is a binary variable indicating whether the respondent nominated friends who were not able to be matched within sample (such as friends outside of school).

Table 2A
Descriptive Statistics by Race

	Obs	Mean	Std.	Obs	Mean	Std.	Obs	Mean	Std.
Smoke	37592	0.39	0.49	10938	0.25	0.43	8331	0.34	0.47
Drink	37512	0.56	0.50	10910	0.51	0.50	8314	0.57	0.50
Get Drunk	37365	0.33	0.47	10793	0.22	0.42	8239	0.30	0.46
Exercise	36493	2.38	1.17	10239	2.07	1.24	7606	2.13	1.21
Any Exercise	36493	0.96	0.19	10239	0.91	0.29	7606	0.93	0.26
Male	38476	0.48	0.50	11745	0.41	0.49	9096	0.47	0.50
White	38619	1.00	0.00	11808	0.00	0.00	9162	0.00	0.00
Hispanic	38619	0.00	0.00	11808	0.00	0.00	9162	1.00	0.00
Black	38619	0.00	0.00	11808	1.00	0.00	9162	0.00	0.00
Asian	38619	0.00	0.00	11808	0.00	0.00	9162	0.00	0.00
Live with Mom	38334	0.94	0.23	11470	0.90	0.30	8897	0.92	0.27
Maternal Years of Education	38619	13.61	2.22	11808	13.47	2.12	9162	12.30	2.63
Maternal Caring Scale	38619	4.78	0.59	11808	4.79	0.60	9162	4.77	0.63
Native Born	37813	0.98	0.14	11496	0.96	0.19	8836	0.72	0.45
Grade = 7	38472	0.14	0.34	11729	0.16	0.36	9089	0.13	0.34
Grade = 8	38472	0.14	0.34	11729	0.16	0.37	9089	0.12	0.32
Grade = 9	38472	0.21	0.41	11729	0.20	0.40	9089	0.22	0.41
Grade = 10	38472	0.20	0.40	11729	0.18	0.39	9089	0.21	0.41
Grade = 11	38472	0.17	0.38	11729	0.16	0.36	9089	0.17	0.37
Grade = 12	38472	0.15	0.36	11729	0.14	0.34	9089	0.15	0.36
Missing	38619	0.35	0.48	11808	0.51	0.50	9162	0.57	0.49
Number of Nominations	38619	2.58	1.52	11808	2.18	1.52	9162	2.03	1.49
Proportion White	34543	0.84	0.27	9844	0.09	0.24	7504	0.27	0.39
Proportion Black	34543	0.03	0.13	9844	0.81	0.33	7504	0.07	0.20
Proportion Hispanic	34543	0.06	0.18	9844	0.05	0.17	7504	0.58	0.44
Proportion Asian	34543	0.03	0.12	9844	0.02	0.10	7504	0.05	0.18
Proportion Other Race	34543	0.04	0.13	9844	0.04	0.14	7504	0.04	0.14
Proportion Mom Less High School	28837	0.12	0.25	7260	0.14	0.28	5266	0.33	0.40
Proportion Mom Some College	28837	0.18	0.27	7260	0.21	0.32	5266	0.14	0.27
Proportion of Mom College Grad	38619	0.36	0.31	11808	0.33	0.29	9162	0.28	0.27
Proportion Native	33645	0.97	0.13	9455	0.96	0.16	7204	0.76	0.37
Proportion Live with Mom	33945	0.94	0.16	9379	0.91	0.21	7218	0.93	0.20

Table 3A
Predictors of Nominating Zero Same Sex/Same Grade Friends

Outcome	No Same Grade/ Gender Friends	No Same Grade/ Gender Friends
Male	0.032***	0.032***
	(0.004)	(0.004)
Hispanic	-0.004	-0.004
	(0.005)	(0.005)
Black	0.009*	0.009*
	(0.005)	(0.005)
Asian	-0.040***	-0.040***
	(0.008)	(0.008)
Maternal Education	-0.002***	-0.002***
	(0.001)	(0.001)
Maternal Caring Index	-0.010***	-0.010***
	(0.002)	(0.002)
Native Born	-0.008	-0.008
	(0.007)	(0.007)
Grade = 8	0.024***	0.016**
	(0.006)	(0.007)
Grade = 9	0.043***	0.027**
	(0.012)	(0.012)
Grade = 10	0.086***	0.066***
	(0.012)	(0.014)
Grade = 11	0.106***	0.085***
	(0.012)	(0.014)
Grade = 12	0.094***	0.071***
	(0.013)	(0.016)
Missing	0.289***	0.289***
	(0.010)	(0.010)
% Black	0.010	0.016
	(0.039)	(0.040)
% Hispanic	-0.121**	-0.109*
	(0.060)	(0.059)
% Mom College Grad	-0.065	-0.040
	(0.042)	(0.041)
% Smoke	-0.010	
	(0.034)	
% Drink		0.066**
		(0.032)
Constant	0.050*	0.015
	(0.026)	(0.026)
Observations	60698	60562
R-squared	0.225	0.225

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4A
Predictors of "Unusual Type"—Single Cluster Membership

Outcome	Single Cluster	Single Cluster	Single Cluster	Single Cluster	Single Cluster	Single Cluster	Single Cluster	Single Cluster
Cluster	Cluster	Cluster	Cluster	Cluster	Cohort-Cluster	Cohort-Cluster	Cohort-Cluster	Cohort-Cluster
Fixed Effects	None	None	School	School-Cluster	None	None	School	School-Cluster
Male	0.004	0.004	0.005*	0.005*	-0.019***	-0.019***	-0.018***	-0.020***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)	(0.006)	(0.006)
Hispanic	0.051***	0.051***	0.019***	0.020***	0.018	0.018	0.008	0.007
	(0.007)	(0.007)	(0.005)	(0.005)	(0.019)	(0.020)	(0.011)	(0.012)
Black	0.032***	0.030***	0.014*	0.015**	0.024	0.019	0.028	0.029
	(0.006)	(0.006)	(0.007)	(0.007)	(0.019)	(0.019)	(0.023)	(0.022)
Asian	0.123***	0.122***	0.078***	0.078***	0.146***	0.142***	0.097***	0.097***
	(0.014)	(0.014)	(0.012)	(0.012)	(0.022)	(0.023)	(0.018)	(0.018)
Maternal Education	0.001*	0.001	0.000	0.000	0.000	-0.000	0.001	0.001
	(0.001)	(0.001)	(0.000)	(0.000)	(0.002)	(0.002)	(0.001)	(0.001)
Maternal Caring Index	0.003*	0.002	0.001	0.001	0.009**	0.006*	0.004	0.004
	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
Native Born	-0.029***	-0.028***	-0.012*	-0.010*	-0.003	0.001	0.015	0.016*
	(0.010)	(0.010)	(0.006)	(0.006)	(0.021)	(0.022)	(0.009)	(0.009)
Grade = 8	-0.013	-0.013	-0.013	-0.004	-0.028**	-0.024*	-0.022*	0.008
	(0.008)	(0.009)	(0.009)	(0.008)	(0.014)	(0.014)	(0.013)	(0.017)
Grade = 9	-0.033***	-0.031***	-0.047**	-0.003	-0.100***	-0.093***	-0.058**	0.030
	(0.012)	(0.012)	(0.022)	(0.008)	(0.019)	(0.020)	(0.023)	(0.024)
Grade = 10	-0.057***	-0.055***	-0.070***	-0.022**	-0.168***	-0.160***	-0.122***	-0.026
	(0.012)	(0.012)	(0.023)	(0.009)	(0.021)	(0.021)	(0.023)	(0.025)
Grade = 11	-0.059***	-0.056***	-0.072***	-0.021**	-0.186***	-0.177***	-0.140***	-0.035
	(0.012)	(0.013)	(0.023)	(0.009)	(0.020)	(0.020)	(0.024)	(0.024)
Grade = 12	-0.057***	-0.054***	-0.070***	-0.020**	-0.171***	-0.161***	-0.124***	-0.016
	(0.012)	(0.012)	(0.023)	(0.009)	(0.019)	(0.020)	(0.022)	(0.025)
Missing	-0.008***	-0.008***	-0.011***	-0.012***	-0.098***	-0.097***	-0.100***	-0.100***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)	(0.008)	(0.007)	(0.007)
Smoke		-0.010***	-0.010***	-0.010***		-0.034***	-0.034***	-0.034***
		(0.002)	(0.002)	(0.002)		(0.005)	(0.005)	(0.005)
Drink		-0.001	-0.003	-0.003		-0.005	-0.008*	-0.009*
		(0.003)	(0.003)	(0.003)		(0.005)	(0.005)	(0.004)
Constant	0.082***	0.089***	0.113***	0.076***	0.490***	0.514***	0.476***	0.400***
	(0.022)	(0.022)	(0.021)	(0.014)	(0.039)	(0.039)	(0.030)	(0.028)
Observations	60706	60392	60392	59988	60706	60392	60392	59988
R-squared	0.029	0.029	0.068	0.081	0.032	0.033	0.072	0.084

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5A
Analysis of the Change in Composition of the Sample Due to Singleton Clusters

Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Group	None	School	Cluster	School/Cluster	School-Cluster	School-Cluster-Cohort	School-Cluster-Cohort-Xs
% Smoke	0.387***	0.367***	0.371***	0.350***	0.309***	0.298***	0.289***
	(0.010)	(0.011)	(0.011)	(0.012)	(0.019)	(0.022)	(0.026)
Observations	50959	50959	50959	50959	50959	50959	50959
R-squared	0.135	0.143	0.245	0.252	0.580	0.649	0.761
Outcome	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke	Smoke
Group	None	School	Cluster	School/Cluster	School-Cluster	School-Cluster-Cohort	School-Cluster-Cohort-Xs
% Smoke	0.387***	0.367***	0.385***	0.365***	0.368***	0.361***	0.353***
	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	50959	50959	47500	47500	32461	28952	21159
R-squared	0.135	0.143	0.136	0.144	0.136	0.135	0.136

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Each column and row is from a separate regression. The first row repeats the results from Table 5 and the second row reproduces the Column 1 results with the non-singleton samples.