Peer Effects and Youth Smoking in the European Global Youth Tobacco Survey

Silda Nikaj[1]

Abstract: This paper investigates the effect of peer smoking on individual smoking among youths in 10 countries that participated in the European Global Youth Tobacco Survey (GYTS). I control for endogeneity in school selection and unobserved school-level characteristics through the use of school fixed-effects. I use instrumental variables to address the simultaneity in peer and individual behaviours. Identification arises by comparing students in different classes within the same school. On average, an increase in the share of classmates who smoke by 10 percentage points increases the probability that an individual in that class will smoke by 3 to 6.9 percentage points. The results imply that any policy intervention such as anti-smoking messages, smoking bans, or higher cigarette prices will be even more cost-effective because of the social multiplier effect of peers – policies affecting some individuals in a group will generate spillovers to others through the peer effect.

Key Words: adolescent smoking, European youths, peer effects, substance use

JEL Codes: I12; D1

Introduction

Smoking comprises the most pressing public health problem of the 21st century (WHO, 2012). Globally, approximately 6 million people die each year from smoking-related disease (Tobacco Atlas, 2012; Jha et al., 2006). Youth smoking is of particular interest because most lifelong cigarette users start smoking in adolescence (Surgeon General, 2012; Chen and Millar, 1998). In recent years the focus of academic research has expanded to explore the social determinants of health behaviours. Previous literature recognizes that economic choices, including health chocies, are made within a social context (Becker & Murphy, 2000; Akerlof, 1997). This is especially true among young people, who are believed to be more influenced by their peers and more likely to participate in risky health behaviours (Powell et. al. 2005; Gaviria and Raphael, 2001, Clark and Loheac, 2007; Lundborg, 2006; Fletcher, 2010; McVicar, 2011).

---

[1] Silda Nikaj, Economist, University of Illinois at Chicago, 610 S. Morgan St. 725 University Hall (MC 144) Chicago IL 60607.

Theoretically, social contexts can alter the benefits and costs of participating in risky behaviours. For example, Becker and Murphy (2000) argue that individuals may experience a higher payoff to a behaviour if social groups endorse that behaviour. This would imply that a larger group of smokers in one's social circle changes the marginal utility of smoking for the individual, and therefore increases the demand for cigarettes (Becker & Murphy, 2000). Alternatively, since social networks are organized under a system of rewards and punishments, social norming implies that individuals do not deviate from the average behaviour of the group because they fear social sanctions (Akerlof, 1997). For example, young people may find it easier to preserve their social standing within a network if they smoke. Finally, individuals mimic behaviour they observe within their own social group because of limited information or knowledge of the particular commodity. Group behaviour then "signals" appropriate behaviour (Bikhchandani et al., 1992).

In this paper I estimate peer effects in smoking in 10 countries that participated in the European Global Youth Tobacco Survey (GYTS). A growing literature in the US and several studies from Europe have examined the role of peer influence in smoking (Ali and Dwyer, 2009; Powell et. al. 2005; Gaviria and Raphael, 2001; Clark and Loheac, 2007; Fletcher, 2010; Pertold, 2009; McVicar, 2011). Nonetheless, little is known about the effect of peer influences on smoking in most countries. In this paper I examine a group of countries where the evidence regarding the social context and smoking is limited. In doing so, I follow the current literature and address biases that arise from school selection, and the simultaneity between peer and individual behaviours. I control for endogeneity in school selection and unobserved school-level characteristics through the use of school fixed-effects. I use instrumental variables to address the simultaneity in peer and individual behaviours. Identification arises by comparing students in different classes within the same school. Consistent with previous literature I find that peer smoking has significant effects on individual smoking among middle and high school students in 10 countries that participated in the European Global Youth Tobacco Survey (GYTS).

1. Background

A burgeoning literature in the US has found that peer effects are important determinants of smoking among youth (Ali and Dwyer, 2009; Powell et. al. 2005; Gaviria and Raphael, 2001; Clark and Loheac, 2007; Fletcher, 2010). A 10 percentage point increase in peer smoking increases youth smoking between 0.9 to 5.8 percentage points. Evidence from Europe has also found peer influence to be one of the primary drivers of youth smoking, though the evidence is more limited here (Lundborg, 2006; McVicar, 2011; Pertold, 2009).

There are several challenges in estimating peer influences on individual behaviours. Methodologically, peer influence is generally confounded by the nonrandom selection of peers and by unobservable peer characteristics. Manski (1993) identifies three effects that could lead to uniform behaviours in a group: endogenous effects, contextual effects, and correlated effects. The *endogenous effect* implies that individual behaviour responds to the behaviour of the peer group. For example, an individual is more likely to smoke if there is a high prevalence of smoking in his/her peer group. *Contextual effects* imply that individuals respond to the exogenous characteristics of the peer group. For example, this effect would imply that the marital status, wealth, and educational attainment of a peer's parents will

affect whether an smokes. Finally, *correlated effects* imply that sorting into schools should produce uniformity in behaviour as individuals that attend the same school are plausibly similar across an array of unobserved characteristics. This may lead not only to correlation in their socio-economic attributes but also in their health behaviours. Failure to control for correlated and contextual effects will bias estimates of peer behaviour influence on individual behaviour (the endogenous effect). Distinguishing between these three effects is important because each effect has different implications for the potential impact of policy interventions. Policy interventions that affect endogenous interaction generate a social multiplier, implying that affecting some individuals in the group though traditional policy instruments (i.e. anti-smoking messages, smoking bans, higher cigarette prices, etc.) will generate spillovers to other group members through the peer effect. Correlated and contextual effects do not generate multiplier effects (Ali and Dwyer, 2009; Fletcher, 2010; Manski, 2000).

Even when researchers do appropriately control for correlated and contextual effects, the endogenous effect is difficult to identify because of the simultaneity between peer and individual behaviour. Peer smoking and individual smoking are jointly determined within a group; it is difficult to separate whether the group affects the individual or the other way around. This simultaneity in behaviour, which Manski calls the "reflection problem", would bias OLS estimates of peer influence.

To address the correlated effects problem I follow the current literature and control for school fixed-effects (McVicar and Polanski, 2014; Clark and Loheac, 2007; Lundborg, 2006; Ali and Dwyer, 2009; Fletcher, 2010; Pertold, 2009). School fixed-effects account for unobservable school-level characteristics that are common to all students in a school, such as school smoking policies, school resources spent on tobacco control programmes, and common background characteristics at the school-level.

In this analysis the peer smoking variable is defined within school and class. The relevant group of peers would be school peers in the same class as the individual concerned. Identification arises because of variation in the share of peers who smoke between classes within schools. There are distinct advantages and disadvantages to defining the peer group too widely or narrowly. For example, a peer group defined to include all students in the school includes a wide range of influences on adolescent behaviours, but also includes distal relational ties that may not be as influential. At the same time, defining the peer group as nominated friends, which unambiguously identifies the most influential group, brings about further selection concerns and estimation bias, as individuals are allowed to choose their own peers. I follow Lundborg (2006) and construct the peer group to include peers in the same class. The advantage of defining the peer group at the class level is that the children in the class continuously interact with one another.

I employ an instrumental variables methodology to address the simultaneity between peer and individual behaviour. In this paper I follow previous literature and use: (1) peers' parental smoking, (2) the frequency with which peers see non-family members smoking in their homes, and (3) random variation in peer age and sex composition to instrument for peer smoking (Ali and Dwyer, 2010; Fletcher, 2010; Bifulco et al., 2011). Parental smoking increases an individual's propensity to smoke. The proportion of peers whose parents smoke has no direct effect on an individual's smoking choices. Similarly, the frequency

with which an individual sees non-family members smoke in their home should affect that individual, but the percentage of peers who report seeing people smoke in their home frequently should have no direct effect on an individual's smoking choices. Here I exploit the Gaviria and Raphael argument (2001) that, conditional on school selection, peer background factors are unlikely to be directly related to individual behaviour in the school setting, but may be related indirectly through peer behaviour. While these instruments would be problematic if the peer group was defined as nominated friends, who may have knowledge of the household environment that the student lives in, peer background characteristics or peer contextual factors may be less problematic when peers are defined at the school or classroom level (Gaviria and Raphael, 2001; Lundborg, 2006). It is difficult to imagine that students have intimate knowledge of the home environment of all the other 30 students in their class. Therefore like previous literature I assume contextual effects are zero (Lundborg, 2006), and use peer average household characteristic as instruments.

I augment these instruments with two additional instruments that account for random variation in the age and sex composition in the classroom. Bifulco and colleagues (2011) find that plausible random variation within school cohorts in the peer racial or peer maternal education composition is uncorrelated with drinking or smoking for the individual. This is in spite of the fact that a student's own race and parental education correlate with their own drinking and smoking behaviour. Like Bifulco and colleagues (2011), I argue that conditional on school fixed-effects, the variation in age and gender composition from class to class within schools is more or less random. Peer age and sex composition should not affect individual smoking directly, except through the fact that older and male peers are more likely to smoke.

There are three papers that are closely related to the current work. Lundborg (2006) uses data for adolescents aged 12 to 18 from Sweden and a school fixed-effects methodology and defines the peer group as peers in the same class. The paper uses the proportion of students living in a single-parent household, the proportion living in an apartment, the proportion with parents born outside Sweden, and the proportion of peers who had drunk alcohol together with their parents as instruments. Lundborg (2006) finds that a 10 percentage point increase in peer smoking increases individual smoking by 4.7 percentage points, an effect equivalent to a 32% change in individual smoking behaviour.

Pertold (2009) examines peer effects and smoking among young people in the Czech Republic. The peer variable is defined as the percentage of peers who are current daily smokers, and the instrument is peers' pre-secondary school daily smoking. The motivation behind the instrument is that pre-existing behaviours, before a student enrolled in secondary school, are not affected by their current peers, and conditional on school fixed-effects should reduce the bias present in naïve estimates of using current peer smoking status. The analysis finds that increasing the share of peers who are daily smokers by 10 percentage points does not affect females but increases smoking among male students by 2.8 percentage points.

A third paper by McVicar (2011) estimates peer effects for smoking in 26 European countries. He utilizes the 2007 wave of the European Schools Survey Project on Alcohol and Other Drugs (ESPAD) and defines the relevant peer group as classmates.

His instruments include the proportion of peers with an older sibling and the proportion of peers living with two parents. One drawback of the data is that most schools only surveyed one class, which precludes the utilization of school fixed-effects in the estimation, and raises concerns that the estimates may suffer from selection bias. Despite this limitation McVicar's paper is able to generate estimates for 15 out of the 26 countries, where the instruments perform well. Overall, the paper finds wide variation in peer effect estimates, ranging from 0.16 in Belgium to 0.591 in Denmark.

The current paper contributes to the discussion of peer effects on smoking in Europe and extends the current literature in several ways. First, I am able to control for sorting into schools and remove bias driven by school selection in the estimates. Lundborg (2006) finds that controlling for school fixed-effects leads to a 20% reduction compared with naïve OLS estimates of peer effects. In my paper, I find estimates that include school fixed-effects to be between 20-50% smaller than those that do not include school fixed-effects. Second, I test whether random cohort variation in the sex and age composition of classes can be used as instruments in identifying the peer effect and find evidence that such use is appropriate. Third, I am able to produce peer effects for some of the areas with the highest prevalence of smoking both in Europe and globally. The analysis includes data from the Czech Republic, Estonia, Latvia, Russia, Bulgaria, Slovakia, and Slovenia, where well over 25% of young people report having smoked at least one cigarette in the past month.

The analysis has important implications for the interactions of peer influence and public policy. I find estimates of peer influence in the 0.2-0.69 range, suggesting that tobacco control interventions, in addition to modifying individual behaviour, will also generate large spillovers into the individual's peer group through the peer effect.

The rest of the paper is organized as follows: section 3 discusses the data and variable construction; section 4 discusses the empirical methodology; sections 5 and 6 present results and discuss the findings.

2. Data Sources and Variable Construction

I use cross sectional data from the European Global Youth Tobacco Survey (GYTS), a school-based survey which examines youth cigarette use, knowledge and attitudes, media exposure, and access to cigarette products among middle and highs school students. GYTS uses a standardized methodology for constructing sampling frames, selecting schools and classes, preparing questionnaires, conducting field procedures, and processing data. GYTS produces samples of grades associated with students aged 13-15 years, although the actual age range in the sample is from 11-19 years. The survey includes all schools, both public and private, in a geographically defined area. All students in selected classes attending school on the day the survey is administered are eligible to participate. Student participation is voluntary and anonymous using self-administered data collection procedures.

The GYTS surveyed 33 European countries between 1999 and 2012. In several countries, the GYTS was administered in multiple waves and years. The data is publicly available on

the Centers for Disease Control and Prevention (CDC) website.[2] The current analysis includes restricted data on 13 of the 33 European countries. The restricted data were provided by CDC with additional variables.[3] The current data set includes both a school identifier and a class identifier, which are missing in the public release data. This analysis could not be conducted without these identifiers. I estimate results for 10 of the 13 countries, because one of the variables, the availability of pocket money, was not included in the survey after 2007. For example, the 2008 surveys of Bulgaria, Kyrgyzstan, Moldova, Serbia, and Uzbekistan were excluded from the analysis because these data did not include the pocket money variable. The final sample upon which the analysis is performed includes 10 countries and approximately 53,000 observations. A summary of the countries and years in the sample is shown in Table 1.

Table 1 Countries Included in the Analysis

Country	Years	Schools	Classes	Observations
Bulgaria	2002	15	35	697
Cyprus	2005	119	780	15,787
Czech Republic	2002, 2007	99	397	6,620
Estonia	2007	45	148	2,704
Kazakhstan	2004	105	577	10,728
Latvia	2007	49	179	3,090
Moldova	2004	60	218	4,105
Russia	2002	20	70	1,403
Slovakia	2003	59	232	4,106
Slovenia	2003	57	218	4,141
Total		628	2,854	53,381

The analysis variables are defined in Table A1, provided in Appendix A. The outcome variable in this paper is smoking participation. Smoking participation is a binary variable that takes a value of 1 if the individual has smoked any cigarettes in the last 30 days and zero otherwise. Summary statistics are provided in Table 2. The data include a group of countries that are well known for high smoking prevalence among young people: Czech Republic, Estonia, Latvia, Russia, Bulgaria, Slovakia, and Slovenia. In these countries, between a quarter and a third of youths report having smoked in the past 30 days. The lowest prevalence is found in Kazakhstan, where approximately 9.4% of young people have smoked cigarettes in the past month. These results closely mirror findings from other global tobacco surveys.

---

[2] GYTS Data Portal: https://nccd.cdc.gov/gtssdata/Ancillary/DataReports.aspx?CAID=2
[3] CDC provided data with school and class identifiers on the following countries: Bulgaria, Cyprus, Czech Republic, Estonia, Kazakhstan, Kyrgyzstan, Latvia, Moldova, Russia, Serbia, Slovakia, Slovenia, and Uzbekistan.

Table 2 Summary Statistics

Country	Smoking Prevalence	Peer Smoking	Male	Age	Only Mother Smokes	Only Father Smokes	Both Parents Smoke	Exposure to Smoking at Home*
Bulgaria	0.280	0.283	0.501	13.680	0.166	0.186	0.415	3.842
	(0.449)	(0.188)	(0.500)	(0.748)	(0.373)	(0.390)	(0.493)	(3.028)
Cyprus	0.152	0.152	0.477	14.138	0.054	0.366	0.134	3.103
	(0.359)	(0.154)	(0.499)	(1.261)	(0.226)	(0.482)	(0.341)	(2.305)
Czech Republic	0.314	0.316	0.510	13.806	0.097	0.211	0.201	1.912
	(0.464)	(0.192)	(0.500)	(1.009)	(0.296)	(0.408)	(0.401)	(2.753)
Estonia	0.307	0.308	0.493	14.659	0.105	0.269	0.145	3.143
	(0.461)	(0.164)	(0.500)	(1.072)	(0.306)	(0.443)	(0.352)	(2.641)
Kazakhstan	0.094	0.097	0.477	13.795	0.029	0.434	0.058	1.831
	(0.292)	(0.137)	(0.499)	(1.141)	(0.167)	(0.496)	(0.234)	(2.046)
Latvia	0.368	0.372	0.453	14.432	0.068	0.320	0.153	2.868
	(0.482)	(0.214)	(0.498)	(1.252)	(0.251)	(0.467)	(0.360)	(3.030)
Moldova	0.162	0.166	0.450	14.002	0.014	0.446	0.045	1.462
	(0.369)	(0.152)	(0.498)	(1.114)	(0.119)	(0.497)	(0.207)	(2.007)
Russia	0.326	0.328	0.490	13.978	0.044	0.411	0.116	1.901
	(0.469)	(0.205)	(0.500)	(1.264)	(0.205)	(0.492)	(0.321)	(2.833)
Slovakia	0.236	0.239	0.497	13.508	0.110	0.234	0.197	2.487
	(0.425)	(0.163)	(0.500)	(1.026)	(0.313)	(0.423)	(0.397)	(2.487)
Slovenia	0.258	0.265	0.441	14.297	0.139	0.175	0.165	3.100
	(0.438)	(0.202)	(0.497)	(1.118)	(0.346)	(0.380)	(0.371)	(3.055)

*Note: Means and Standard Deviations. * Days in the past week exposed to smoking at home from a non-family member.*

Our variable of interest is the peer smoking measure, which is constructed as the percentage of students in the class who smoke cigarettes, excluding the individual in question. Individual explanatory variables include age measured in years, sex defined as a dichotomous variable (1 if male, zero otherwise), and pocket money expressed in the local currency. In almost all countries the average age of survey participants is 14 and about half of the sample is male. I also have information on family smoking. Parental smoking is defined as a dichotomous variable, taking a value of 1 if any of the parents smoke and 0 if none smoke. In the table the variable is further broken down according to whether only the mother smokes, only the father smokes, or both parents smoke. A large proportion of parents smoke and students are exposed to smoking from non-family members at home between 1.5 to 4 days a week, suggesting that these young people live in an environment where smoking is commonplace.

3. Empirical Specification and Identifying Assumptions

The framework below specifies a model of school-class-based peer effects on the probability that a student i in class c of school s will smoke. S_{ics} is a dichotomous variable indicating smoking participation in the last 30 days.

$$S_{ics} = \beta_0 + \beta_1 P_{ics} + \beta_2 X_{ics} + \beta_3 F_{ics} + \beta_4 A_{ics} + I_s + \varepsilon_{ics} \qquad (1)$$

$P_{ics,}$ the peer effect measure, is the percentage of students who smoke within student i's class, excluding student i him/herself.[4] X_{ics} is a vector of personal characteristics (age, sex, pocket money). Pocket money is comprised of several categories, capturing the amount of available spending money the youths have each month. The pocket money variable is included as a series of dummies, with the lowest category of zero allowance being the excluded category.

F_{ics} includes information on parental smoking. All parental smoking variables are defined as binary variables. The excluded category is where no parents smoke. A_{ics} are age dummies, which account for the fact that individuals are more likely to report smoking as they age. The lowest age group in the sample is the omitted benchmark.

School dummies are indicated by I_s. School dummies are intended to capture the effect of students sorting into schools, as well as school-level variables that all students are exposed to (i.e. anti-smoking sentiments, smoking bans on school grounds, common cigarette prices etc.). The utilization of school fixed-effects forces identification of the peer effect by comparing students in different classes within the same school and requires that there be variation within schools in the peer variable. Therefore each school in the analysis must have data for at least two classes. The number of classes per school in the sample ranges between 2 and 15.[5]

3.1. Instruments and Identifying Assumptions

As previously stated, one must account for the fact that the individual can affect the peers' behaviour, while at the same time the peers affect his/her own behaviour - the "reflection problem". This simultaneity in behaviours generates bias in OLS estimates of β_1. Instrumental variables (IV) can produce consistent estimates of peer influence on individual behaviour. Identifying appropriate instruments (Z) requires that the instruments are correlated with peer smoking and are uncorrelated with determinants of individual smoking. In our case, these assumptions would mean that the instruments (Z_{ics}) have a clear effect on the peer smoking variable (P_{ics}), which implies that $\alpha_1 \neq 0$ in the first stage equation in the two stage least squares (2SLS) framework below (Equation 2). The second assumption, often referred to as the exclusion restriction(s), requires that the only

---

[4] The peer variable within class and school is defined as the share of students who smoke, excluding the individual, using the following formula $P_{ics} = \frac{1}{n_{cs}-1}\sum_{j \neq i} S_{jcs}$, where n_{cs} is the number of students in the class, S_{jcs} is the binary smoking variable, and $j \neq i$.

[5] I did not want to exclude observations where there were a small number of classes (2) as this would generate bias in the estimates. I realize this strategy makes the standard errors larger because it has less variation, but it seems reasonable to trade off loss of efficiency to reduce bias.

pathway through which the instruments may affect individual smoking (S_{ics}) is through the effect the instruments exert on the peer variable (P_{ics}). A second part of this assumption implies that conditional on covariates – all individual and school level controls – the instrument is independent of potential outcomes. In essence, the second set of assumptions implies that $Cov(Z_{ics}, \varepsilon_{ics}) = 0$, or that the instruments are uncorrelated with any observed or unobserved determinants of individual smoking.

$$P_{ics} = \alpha_0 + \alpha_1 Z_{ics} + \alpha_2 X_{ics} + \alpha_3 F_{ics} + \alpha_4 A_{ics} + I_s + \nu_{ics} \qquad (2)$$

$$S_{ics} = \beta_0 + \beta_1 \hat{P}_{ics} + \beta_2 X_{ics} + \beta_3 F_{ics} + \beta_4 A_{ics} + I_s + \varepsilon_{ics} \qquad (3)$$

My analysis in this paper uses the proportion of class peers whose parents smoke, the frequency with which class peers see non-family members smoking at home, and the age and sex composition of the class peers to instrument for peer smoking. The first two instruments take advantage of the fact that the individual has limited knowledge of their peers' home environments. Ali and Dwyer (2009) and Fletcher (2010) utilize instruments similar to those I use here. Ali and Dwyer (2009) employ the proportion of peers whose parents smoke as an instrument. The intuition behind the instruments is that although classmates whose parents smoke are themselves more likely to smoke, the proportion of a student's classmates whose parents smoke should have no direct effect on that student's smoking behaviour. In other words, the only way that my classmates' parental smoking may affect my smoking is through increasing the propensity that my classmates smoke. A similar argument can be made for the frequency that peers observe non-family members smoking in their home environment. A higher frequency of smoking at peers' homes may increase peer smoking but should have no direct effect on individual smoking except through the channel of increasing peer smoking. Both these two instruments rely on the fact that classmates' home environments are unobserved to the individual. This assumption is more likely to be satisfied when the peer group is defined at the school or classroom level (Lundborg, 2006).

For the additional instruments, I rely on evidence from the US literature on the health outcomes of young people to identify instruments. More particularly, Bifulco and colleagues (2011) find that, conditional on observables and school fixed-effects, plausible random variation within school cohorts in the peer racial or peer maternal education composition is uncorrelated with contemporaneous smoking for the individual. This is in spite of the fact that a student's own race and parental education correlate with their own drinking and smoking.

The second set of instruments relies on random variation in the age and sex composition of class peers. Ex ante we know that males and older adolescents are more likely to smoke than females and younger adolescents – this is true in our data. I argue that conditional on school fixed-effects and observables, the variation in age and gender composition from class to class within schools is more or less random. Some students end up by chance in a class with more males, or a class with a slightly older group of peers. Peer age and sex composition should not affect individual smoking directly, except through the fact that if an adolescent is placed in a classroom with more older and male peers she/he is more likely to be exposed to a higher share of smokers in his/her class peer group. I provide evidence below that these two instruments perform

well – they do predict peer smoking and are uncorrelated with the second stage error term.

First, I estimate a linear probability model (LPM) using ordinary least squares (OLS); I then use two-stage least squares (2SLS) to address the endogeneity of the peer measure. I cluster the standard errors at the school level to allow for errors to be correlated within the school. I conduct the analysis at the country level, because pooling the results into one large sample would assume that peer effects are the same across all countries. Moreover, since most tobacco control programmes are conducted at the country level, country-level estimates may be more relevant from a policy perspective.

4. Results

4.1. OLS Estimates

Table 3 summarizes the OLS results. An increase in peer smoking is positively and significantly associated with a higher probability of individual smoking in all countries. The point estimates range between .192 in Estonia and .445 in Moldova. These estimates suggest that a 10 percentage point increase in peer smoking is associated with an increase in individual smoking of between 1.9 to 4.45 percentage points. The largest relative effect is present in Kazakhstan. Kazakhstan has a smoking prevalence of 9.4%, thus a 4.15 percentage point effect suggests an almost 45% change in individual smoking from a 10 percentage point change in peer smoking. The smallest relative effects are present in the Czech Republic and Estonia where a 10 percentage point increase in peer smoking is associated with 9% and 6% increases in individual smoking, respectively.

The other controls have the expected sign. Parental smoking is associated with a higher probability that the individual in question smokes. Maternal smoking and smoking by both parents appear to exert the largest impact. Smoking by fathers remains an important determinant of youth smoking, but its relative impact is smaller. Exposure to smoking by non-family members at home is associated with an increased likelihood of individual smoking. These findings are not surprising: where parents are more permissive about smoking at home, the youth may interpret this behaviour as a licence to smoke. Though not shown, smoking participation rises with age and with the increased availability of pocket money. Males are more likely to smoke than females. Overall, naïve OLS estimates suggest that higher smoking prevalence among class peers is associated with higher smoking prevalence for the individual. Moreover, these OLS estimates are in line with findings from previous studies in Europe (Lundborg, 2006; McVicar, 2011; Pertold, 2009).

4.2. 2SLS Estimates

Results from the 2SLS estimation can be found in Table 4. The estimates for other controls do not differ between IV and OLS and therefore are not shown in Table 4. I present the results of the peer effect along with F-statistics from the first stage regression – the regression of peer influence on all controls from Table 3 and the instruments. The instruments performed well in the sample for Cyprus, the Czech

Republic, Estonia, Kazakhstan, Latvia, Russia, Slovakia and Slovenia. For all first stage regression the F-statistics are well above the cutoff of 10 suggested by Stock and Staiger (1997). In addition, over-identification tests summarized by the Hansen J Statistic fail to reject the null of exogenous instruments. This implies that the instruments are correctly excluded from the second stage regression. The instruments did not perform well in the case of Bulgaria, and Moldova, where they were not strong enough in predicting the peer variable.

In the Czech Republic a 10 percentage point increase in peer smoking is linked to a 2.95 percentage point increase in smoking, or approximately a 9% increase in smoking prevalence for the individual. In Latvia, a similar 10 percentage point increase in peer smoking is linked to a 2.6 percentage point increase in smoking for the individual – or a 7% effect. The 2SLS point estimate for Estonia is similar to the OLS estimate. A 10 percentage point increase in peer smoking increases individual smoking by approximately 2 percentage points. The results for Slovakia are insignificant, in part because 2SLS produces larger standard errors than OLS. The 2SLS point estimate suggests an effect of 0.25, which is similar in magnitude to what was found in the OLS estimates, and similar to the results for the neighbouring Czech Republic and Baltic states.

The biggest changes in the estimates are recorded in Cyprus, Kazakhstan, Russia, and Slovenia, with an increase of between one third and one half. In Slovenia a 10 percentage point increase in peer smoking increases individual smoking by 4.3 percentage points. In Cyprus, Kazakhstan and Russia a similar increase in peer smoking is linked to increases in individual smoking of 5.3, 6.4 and 6.9 percentage points, respectively. All these estimates highlight the important influence that peers exert on young people's smoking behaviours.

4.3. Robustness and Identification Checks

To determine whether the instruments were appropriately chosen, I subject the data to various robustness checks. First I present the F-statistics and the first stage regressions for each country (Table 5). Stock, Wright and Yogo (2002) suggest that an F-statistic of 10 or greater in first stage results can be reliable in 2SLS estimates. I then test the exclusion restrictions by utilizing Hansen J Test P-values. The Hansen J Test joint null hypothesis states that the instruments are valid instruments, or that the instruments are correctly excluded from the estimated second stage equation. Failure to reject the null, or P-value >0.05, suggests valid instruments. These tests are summarized in Table 5 along with the instruments utilized in each country. Based on the evidence, the first stage is strong for every country except Bulgaria, and Moldova, which both have F-stats of less than 10 from the first stage. I do not interpret results for these countries.

As an alternative test, I take out all other instruments and re-estimate the first stage results using only the strongest predicting instrument for each country. The results are summarized in Table 6 (column 3) and I provide F-statistics of the first stage single instrument results in the 4th column. Removing all but the strongest instrument produces estimates that are not statistically different from those presented in the 2SLS estimates which augment all instruments.

Table 3 Linear Probability Model - OLS Estimates of Peer Smoking on Individual Smoking

Variable	Bulgaria	Cyprus	Czech Republic	Estonia	Kazakhstan	Latvia	Moldova	Russia	Slovakia	Slovenia
Peer Smoking	0.437***	0.337***	0.291***	0.192*	0.415***	0.386***	0.445***	0.370**	0.291***	0.341***
	(0.097)	(0.048)	(0.057)	(0.072)	(0.036)	(0.053)	(0.056)	(0.092)	(0.071)	(0.081)
Male	-0.099*	0.075***	-0.02	-0.002	0.052**	0.057**	0.143***	0.114***	0.037**	-0.016
	(0.050)	(0.010)	(0.013)	(0.019)	(0.007)	(0.021)	(0.017)	(0.027)	(0.015)	(0.017)
Only Mother Smokes	-0.021	0.067***	0.147***	0.089***	0.087**	0.090*	0.073	0.276***	0.055***	0.027
	(0.053)	(0.013)	(0.023)	(0.023)	(0.035)	(0.043)	(0.060)	(0.062)	(0.020)	(0.022)
Only Father Smokes	-0.101*	0.025***	0.099***	0.021	0.009*	0.052***	0.032***	0.053**	0.048***	0.021
	(0.055)	(0.006)	(0.016)	(0.021)	(0.004)	(0.019)	(0.010)	(0.023)	(0.016)	(0.020)
Both Parents Smoke	0.02	0.101***	0.159***	0.087***	0.123***	0.080***	0.055*	0.118**	0.108***	0.074***
	(0.060)	(0.012)	(0.021)	(0.028)	(0.022)	(0.028)	(0.032)	(0.042)	(0.017)	(0.025)
Exposure to Smoking	0.024*	0.010***	0.019***	0.066***	0.006***	0.021***	0.007**	0.009*	0.016***	0.006*
	(0.012)	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
N	697	15,787	6,620	2,704	10,728	3,090	4,105	1,403	4,106	4,141
Mean Smoking	0.280	0.152	0.314	0.307	0.092	0.368	0.162	0.326	0.236	0.258
R²	0.113	0.128	0.101	0.236	0.117	0.164	0.175	0.127	0.135	0.104

All standard errors are clustered at the school level. All regressions include age dummies and dummies for pocket money.

Finally, I run Limited Information Maximum Likelihood (LIML) to check for over-identification of the 2SLS estimates. LIML provides an additional test for weak instruments. LIML is approximately median-unbiased for over-identified constant effects models and provides an alternative to the just-identified tests presented above (Angrist and Pischke, 2009). If 2SLS and LIML estimates come out different, then concerns over weak instruments are valid and alternative instruments should be used. These estimates are summarized in Table 6. The LIML and 2SLS estimates are almost identical to one another, suggesting the instruments are valid.

5. Discussion

This analysis is the first study to estimate the effect of peer influences on individual smoking behaviour among European GYTS participating countries. Overall, the results provide evidence of the importance of peer influences for youth smoking behaviours. I find consistent evidence that youth smoking is responsive to peer influences and that a 10 percentage point increase in peer smoking increases the probability of youth smoking by 2 – 6.9 percentage points. Both the OLS and 2SLS estimates imply that any policy intervention to reduce youth smoking will be even more cost-effective because of the social multiplier effect of peers – policies affecting some individuals in the peer group will generate spillovers for others through the peer effect.

I address several challenges to estimating peer influences at the school level. I use school fixed-effects to address selection into schools and unobserved school characteristics. The inclusion of school fixed-effects reduces estimates of peer effects by as much as 50%.[6] Lundborg (2006) previously found a 20% reduction in estimates when including school fixed-effects. All in all, these findings highlight the importance of removing the variation that arises in group behaviour because of sorting into schools. I then use variation in behaviours between classes within schools to identify the effect of peer smoking on individual smoking.

Overall, the results verify findings from the existing literature in the US and emerging studies from Europe that peer smoking is an important determinant of smoking among young people (McVicar and Polanski, 2014; Clark and Loheac, 2007; Lundborg, 2006; Ali and Dwyer, 2009; Fletcher, 2010; McVicar, 2011; Pertold 2009). Previous studies have found that a 10 percentage point increase in peer smoking is linked to increases in individual smoking of between 1 and 5.8 percentage points. I find estimates of peer effects of between 0.2 to 0.69 percentage points; this result is comparable to those reported in other studies examining peer effects and smoking in Europe. McVicar (2011) estimates effects in the range 0.15 to 0.59. Lundborg (2006) finds a point estimate of approximately 0.47 for Swedish youth. Finally, Pertold (2009) who examines peer effect and smoking by gender among young people in the Czech Republic finds no peer effects for females, but that a 10 percentage point increase in peer smoking increases smoking among male students by 2.8 percentage points. While that result is not fully comparable to those presented here, since the current analysis is not stratified by gender, the effect Pertold (2009) finds is similar to the point estimates found for the Czech Republic in this analysis.

---

[6] Results are available upon request.

I also find country to country differences in the estimated peer effects, from a low of 0.19 in Estonia to a high of 0.69 in Russia. These differences in peer effects imply that policy interventions to reduce youth smoking, such as higher taxes on cigarettes, anti-smoking campaigns, bans on smoking in public and private places, and bans on cigarette advertising, are heterogeneously transmitted through peer influence (Powell et al. 2005; Fletcher, 2010; Cutler and Glaeser, 2010). Where peer influences are larger, such as in Cyprus and Kazakhstan, these policies will be quickly disseminated through the peer effect. However, in places where the peer influence is smaller the interaction between tobacco control policies and social influence will generate smaller multiplier effects, and tobacco control programmes will have to invest differentially to generate large changes in behaviour.

While this paper has been able to address some important issues in estimating peer effects on individual behaviours, some limitations remain. Previous studies were able to control for a rich set of individual characteristics: parental education, family income, racial and ethnic background, and home environment. To the extent that these characteristics are correlated with the outcome and the peer variable, the estimates may reflect the peer effect with some degree of bias.

Moreover, the paper assumes that all sorting takes place at the school level with no additional sorting into classes. Sorting into classes could take place if parents are able to choose the class into which their child is placed - for example, by choosing particular teachers or characteristics of the peer group. However, in most countries, the F-tests and Hansen J-test suggest that sorting based on class peer characteristics is unlikely. Even so, I cannot rule out the possibility that some sorting takes place at the class level.

Finally, my analysis is only conducted for middle and high school students and should be interpreted only as reporting peer effects at the school level. While school peers may represent the most relevant group of peers for youth who are enrolled in school, the analysis should not be interpreted to extend to youths outside of secondary educational institutions.

Disclosure Statement: No potential conflict of interests has been reported by the author.

Table 4 Linear Probability Model - 2SLS Estimates of Peer Smoking on Individual Smoking

Variable		Bulgaria	Cyprus	Czech Republic	Estonia	Kazakhstan	Latvia	Moldova	Russia	Slovakia	Slovenia
Peer Smoking		0.253	0.529***	0.295**	0.190*	0.638***	0.260*	1.075***	0.692***	0.247	0.430***
		(0.551)	(0.049)	(0.117)	(0.099)	(0.096)	(0.152)	(0.211)	(0.108)	(0.155)	(0.142)
Hansen J value	P-	+	0.376	0.276	0.088	0.265	0.907	0.672	0.340	0.742	0.872
First Stage Stat.	F-	3.65	28.096	26.127	40.675	33.847	13.448	8.083	21.247	28.523	15.731
N		697	15,787	6,620	2,704	10,728	3,090	4,105	1,403	4,106	4,141
Mean Smoking		0.28	0.152	0.314	0.307	0.094	0.368	0.162	0.326	0.236	0.258
adj. R²		0.151	0.195	0.151	0.249	0.204	0.222	0.186	0.179	0.16	0.22

All controls from Table 3 are included in the estimation. For additional detail see note in Table 3. +Only one instrument was used in the case of Bulgaria.

Table 5 First Stage of 2SLS Regression - Outcome Peer Smoking

Variable	Bulgaria	Cyprus	Czech Republic	Estonia	Kazakhstan	Latvia	Moldova	Russia	Slovakia	Slovenia
Peer Parental Smoking	0.249***		0.181*		0.242***		0.223*			
	(0.055)		(0.098)		(0.052)		(0.119)			
Frequency Peers Exposed to Smoking at Home	0.043***		0.046***	0.107***		0.052***		.090***	0.048***	0.040**
	(0.007)		(0.015)	(0.013)		(0.016)		(0.023)	(0.016)	(0.015)
Sex Composition	0.165***									
	(0.040)									
Age Composition	.149*		0.069***	0.023*	0.040***	0.056***	0.041***	0.354**	0.080***	0.073***
	(0.077)		(0.010)	(0.014)	(0.006)	(0.015)	(0.011)	(0.137)	(0.011)	(0.027)
F-Statistic	3.65	28.096	26.127	40.675	33.847	13.448	8.083	21.247	28.523	15.731
Hansen J P-Value	-	0.376	0.2763	0.0882	0.2647	0.9066	0.6715	0.34	0.7418	0.8719
N	697	15,787	6.62	2,704	10,728	3.09	4,105	1,403	4,106	4,141
adj. R²	0.499	0.597	0.542	0.562	0.621	0.531	0.491	0.65	0.49	0.685

Table 6 Robustness Checks – Linear Probability Models

	2SLS	LIML	Strongest Instrument Estimates	First Stage F-Stat. Strongest Instrument
	Peer Estimate	Peer Estimate	Peer Estimate	
Bulgaria	0.253	0.253	0.253	3.65
	(0.551)	(0.526)	(0.551)	
Cyprus	0.529***	0.530***	0.587***	42.48
	(0.049)	(0.049)	(0.069)	
Czech Republic	0.295**	0.295**	0.378***	19.35
	(0.117)	(0.116)	(0.137)	
Estonia	0.190*	0.190*	0.209**	73.22
	(0.099)	(0.098)	(0.095)	
Latvia	0.260*	0.260*	0.237	13.12
	(0.152)	(0.150)	(0.255)	
Kazakhstan	0.638***	0.638***	0.536***	25.74
	(0.096)	(0.095)	(0.130)	
Moldova	1.075***	1.076***	0.953***	3.99
	(0.211)	(0.209)	(0.268)	
Russia	0.692***	0.692***	0.797***	14.40
	(0.108)	(0.105)	(0.130)	
Slovakia	0.247	0.247	0.223	55.15
	(0.155)	(0.154)	(0.192)	
Slovenia	0.430***	0.430***	0.407*	12.20
	(0.142)	(0.140)	(0.229)	

References

ALI, M.M. and DWYER, D.S., 2009. Estimating peer effects in adolescent smoking behavior: a longitudinal analysis. *Journal of Adolescent Health*, *45*(4), pp.402-408.

AKERLOF, G.A., 1997. Social distance and social decisions. *Econometrica: Journal of the Econometric Society*, pp.1005-1027.

BECKER, G.S. and Murphy, K.M., 2009. *Social economics: Market behavior in a social environment*. Harvard University Press.

BIFULCO, R., Fletcher, J.M. and Ross, S.L., 2011. The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, *3*(1), pp.25-53.

BIKHCHANDANI, S., HIRSHLEIFER, D. and WELCH, I., 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, *100*(5), pp.992-1026. http://doi.org/10.1086/261849

CHEN, J. and MILLAR, W.J., 1998. Age of smoking initiation: implications for quitting. *Health reports-statistics*

Canada, *9*, pp.39-48.

CLARK, A.E. and LOHÉAC, Y., 2007. "It wasn't me, it was them!" Social influence in risky behavior by adolescents. *Journal of health economics*, *26*(4), pp.763-784. http://doi.org/10.1016/j.jhealeco.2006.11.005

CUTLER, D.M. and GLAESER, E.L., 2007. *Social interactions and smoking* (No. w13477). National Bureau of Economic Research.

FLETCHER, J.M., 2010. Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics*, *19*(4), pp.466-484. http://doi.org/10.1002/hec.1488

GAVIRIA, A. and RAPHAEL, S., 2001. School-based peer effects and juvenile behavior. *The review of economics and statistics*, *83*(2), pp.257-268.

JHA, P., CHALOUPKA, F.J., MOORE, J., GAJALAKSHMI, V., GUPTA, P.C., PECK, R., ASMA, S. and ZATONSKI, W., 2006. Disease control priorities in developing countries. *Disease control priorities in developing countries*.

LUNDBORG, P., 2006. Having the wrong friends? Peer effects in adolescent substance use. *Journal of health economics*, *25*(2), pp.214-233. http://doi.org/10.1016/j.jhealeco.2005.02.001

MANSKI, C.F., 1993. Identification of endogenous social effects: The reflection problem. *The review of economic studies*, *60*(3), pp.531-542.

MANSKI, C.F., 2000. *Economic analysis of social interactions* (No. w7580). National bureau of economic research.

MCVICAR, D. and POLANSKI, A., 2014. Peer effects in UK adolescent substance use: Never mind the classmates?. *Oxford Bulletin of Economics and Statistics*, *76*(4), pp.589-604. http://doi.org/10.1111/obes.12030

MCVICAR, D., 2011. Estimates of peer effects in adolescent smoking across twenty six European countries. *Social Science & Medicine*, *73*(8), pp.1186-1193. http://doi.org/10.1016/j.socscimed.2011.08.006

PERTOLD, F., 2009. Sorting into Secondary Education and Peer Effects in Youth Smoking. Available at: http://cerge-ei.cz/pdf/wp/Wp399.pdf (Accessed: 30th June 2017)

POWELL, L.M., TAURAS, J.A. and ROSS, H., 2005. The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior. *Journal of health Economics*, *24*(5), pp.950-968.

The Tobacco Atlas, World Health Organization, 2012: http://www.tobaccoatlas.org/

World Health Organization, 2012. WHO global report on mortality attributable to tobacco. In *WHO global report on mortality attributable to tobacco*. Available at:. http://www.who.int/tobacco/publications/surveillance/rep_mortality_attributable/en/ (Accessed: 30th June 2017)

Appendix A

Table A1 Definition of Analysis Variables

Variables	Definitions
Smoking Participation	1= Reported smoking cigarettes in the last 30 days, 0 = otherwise
Peer Smoking	Percent of peers within one's grade who report having smoked in the past 30 days, excluding the individual
Age	Respondent's age in years
Male	1 = male, 0 = female
Pocket money	Monthly amount of allowance or pocket money that students receive, measured in local currency.
Only Mother Smokes	1= if mother smokes, 0 = otherwise
Only Father Smokes	1 = if father smokes, 0 otherwise
Both Parents Smoke	1 = if both parents smoke, 0 otherwise
No Parental Smoking	1 = if students report no parental smoking, 0 otherwise. The excluded category
Instruments	
Peer Parental Smoking	Percent of peers with parents who smoke.
Peer Exposure to Smoking at Home	Percent of peers who observe smoking at home, from people other than family.
Age-i	Age composition of peer group
Male-i	Sex composition of peer group