

# Sorting into Secondary Education and Peer Effects in Youth Smoking

Filip Pertold

CERGE-EI\*

## Abstract

The start of daily smoking is often after re-sorting of students between elementary and secondary education. I employ a novel identification strategy based on this re-sorting, in order to estimate peer effects in youth smoking. The reflection problem is addressed by peers' pre-secondary-school smoking, which is not influenced by the current interaction. The self-selection is minimized by own pre-secondary school behavior and the pre-existing smoking prevalence among older schoolmates. The empirical findings from the Czech Republic, where the prevalence of youth smoking has recently reached high levels, suggest that male youth smoking is affected by classmates, while female smoking is not.

***JEL Classification:*** I12;D1

***Keywords:*** Peer effects, Smoking, Sorting

---

\*CERGE-EI is a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of the Academy of Sciences of the Czech Republic.

This research was supported by a grant from the CERGE-EI Foundation under a program of the Global Development Network. This project also benefited from my stay at the University of Illinois in Chicago as a visiting scholar. All opinions expressed are those of the author and have not been endorsed by CERGE-EI, the GDN or by the University of Illinois in Chicago. I would like to thank (in alphabetical order) Alena Bičáková, Ladislav Csémy, Štěpán Jurajda, Robert Kaestner, Peter Katusčák and Lisa Powell for their valuable comments and the National Monitoring Center for Drugs and Drug Addiction for providing me with the Czech ESPAD data. The usual disclaimer applies.

Address: CERGE-EI, Charles University and Academy of Sciences of the Czech Republic, Politických veznu 7, Prague 1, 111 21, Czech Republic. E-mail: [filip.pertold@cerge-ei.cz](mailto:filip.pertold@cerge-ei.cz)

## 1. Introduction

Smoking is a major cause of lung cancer, chronic bronchitis and other diseases (Chaloupka and Warner, 2000). The habit usually starts during secondary-school age, when youth underestimates the health consequences of smoking and the addictive nature of tobacco. The current sociological and economic literature suggests that many youth outcomes can be determined by social interactions (Glynn, 1981). In particular, student smoking can be affected by the smoking of classmates, who can affect the costs of obtaining cigarettes or who provide important information about smoking. Such peer effects can be especially important at a young age. This paper explores the smoking behavior of freshmen at secondary schools approximately seven months after enrollment. Specifically, I test the effect of the new peers on students' daily smoking uptake.

From a public policy perspective the existence of social interactions is important, because they determine the efficiency of government policies designed to affect youth behavior. If group members affect each other, a policy that attempts to influence their attitude towards smoking has two effects – direct and indirect. The direct effect decreases smoking by shifting the norms of smokers. The indirect effect decreases demand even further by the multiplication of behavior as individuals are influenced by their peers. Peer effects thus amplify public policy interventions against smoking.

However, the estimation of peer effects is methodologically complicated, because, as Manski (1995) points out, observed similar behavior in a group does not prove the existence of social interactions within the group. The goal of this paper is to provide an identification strategy that deals with the main identification issues of estimating the effect of peers on smoking uptake, and that is generally applicable in most countries where students are re-sorted between primary and secondary schools.

Manski (1995) defines three possible sources of similar behavior in a group: endogenous, contextual, and correlated effects. The endogenous effect is defined as the effect of peers' behavior on actual individual decisions, while the contextual and correlated effects are confounding factors that can also result in similar behavior within a group, but do not imply the existence of a social multiplier. Specifically, the contextual effect allows for the behavior of a member of a group to be directly influenced not by peers' behavior, but by their characteristics. For example, peers' parents can directly influence individual behavior through restrictions on smoking during a visit at their home. The correlated effect captures other factors that can result in similar behavior and are not related to social interactions with peers. Students may self-select into a group based on similar unobserved preferences toward smoking or their smoking can be affected by an unobserved school-specific anti-smoking policy.

Another important identification issue is the reflection problem (Manski, 1995), which stems from the nature of social interactions – group behavior is always the aggregation of individual behavior and it is difficult to distinguish who influences whom in a peer group. Not addressing the reflection problem causes an upward bias of the estimated peer effects, similarly to the case of self-selection.

We introduce an identification strategy that addresses these key identification issues by using the typical institutional setting of a secondary schooling system combined with information about the initiation of smoking on a daily basis. Specifically, the strategy relies on the re-sorting of peers (classmate groups) between elementary and secondary education in the Czech Republic and on the availability of a survey of smoking behavior of secondary school students.

An important feature of the Czech secondary education system, as in other countries,<sup>1</sup> is the enrollment of students into various secondary schools based on an admission process taking place in the 9<sup>th</sup> grade. Of course, enrollment is driven by individual choice and by entrance exams organized by schools, but to some extent such enrollment into secondary schools is a natural experiment that assigns students to new peer groups. In this paper, this strategy is applied in the Czech Republic using the ESPAD survey, covers not only information on current and pre-secondary school smoking, but also the behavior of an older cohort at the given secondary school.

To reduce the bias caused by the correlated effect, I use several controls for individual pre-secondary school behavior (first cigarette use, consumption of alcohol and marijuana) to predict current daily smoking. In addition, the smoking experience of third-year schoolmates is used to proxy school specific anti-smoking policies or sentiment.

To alleviate the reflection problem as well as the contextual problem I use pre-secondary school classmates' smoking instead of the current smoking of peers as the key explanatory variable. Pre-secondary school smokers are those peers who affect non-daily smokers, who, in turn, make their decision about taking up daily smoking. Based on the re-sorting of classmates, one can define who influences whom in a peer group and hence solve the reflection problem. This approach also identifies the effect of other time constant characteristics of peers on individual smoking uptake, i.e. identify the contextual effects.

In what follows, a peer group is defined as a class at a secondary school. The key outcome variable is current daily smoking of an individual student. This approach follows Lundborg (2006), who defines a smoker as an individual reporting smoking daily or almost

---

<sup>1</sup> Re-sorting between elementary and secondary school is typical for a majority of European nations, except Great Britain and the Scandinavian countries.

daily. The data also allow us to distinguish pre-secondary daily smokers from students who start to smoke daily only in secondary school.

The case of the Czech Republic is an interesting one to study. The proportion of 16-year-old high school students reporting daily smoking is 26% and those reporting having smoked more than 40 cigarettes in their life is 40%, which is among the highest rates in Europe (Figure 1). The high proportion of young smokers suggests that tobacco control policy will have to play an important role in public health policies, and it is hoped that this research will be useful in designing them. This research also provides the first evidence of peer effects among youth in secondary schools with early tracking of children. The previous literature has generally examined peer effects among college students (Kremer and Levy, 2008) or in secondary school systems that do not re-sort students into different schools. The stream of literature that examine peer effects among secondary schools student usually identifies peer effects by using within-school variation in peers' smoking and using a traditional instrumental variable approach to address the reflection problem. My approach is in this context a novel one, because it allows for identification of peer effects in the more traditional institutional set-up and with data that cover only one class in each cohort in a secondary school.

The results suggest that peers do affect individuals' daily smoking at Czech secondary schools with a significant difference between males' and females' smoking behavior: male students are significantly more affected by peers' smoking. These findings are in line with the current literature (e.g. Kremer and Levy, 2008), which finds male students to be more involved in fraternities. Therefore, anti-smoking policies targeted at youth (e.g., smoking bans or information campaigns) can rely on peer effects as a reinforcement mechanism among male students.

## 2. Literature Review and Basic Methodological Issues

The basic econometric specification used for estimating peer effects generally has the following form:

$$(1) \quad \Pr(\text{smoke}_{i(g)}) = \alpha_0 + \alpha_1 \overline{\text{peer}}_{-i(g)} + \alpha_2 X_{i(g)} + \alpha_3 \overline{X}_{-i(g)} + \varepsilon_{i(g)}$$

where  $\Pr(\text{smoke}_{i(g)})$  is the probability of an individual  $i$  in a group  $g$  to be a daily smoker,  $\overline{\text{peer}}_{-i(g)}$  is the average daily smoking of his/her peers in the group (after excluding individual  $i$ ),  $X_{i(g)}$  is the vector of an individual's characteristics,  $\overline{X}_{-i(g)}$  is the vector of average peers' characteristics, and  $\varepsilon_{i(g)}$  is the disturbance.

The three most-often addressed problems encountered when estimating equation (1) are reflection, self-selection, and the omission of an antismoking sentiment. All three bias the estimate of the endogenous effect ( $\alpha_1$ ) upward. The reflection problem is connected to the problem of reverse causality between  $\overline{\text{peer}}_{-i(g),t}$  and  $\Pr(\text{smoke}_{i(g),t})$ , because the researcher cannot observe who influences whom in a class or other peer group.

Finding a solution to the reflection problem is difficult. Kremer and Levy (2008) summarize the recent literature on peer effects among college students and suggest that students' outcomes should be regressed on the pre-college outcomes of their peers rather than on contemporaneous peers' behavior. The important point is that current peers could not affect each others' behavior before they were enrolled in college. Using lagged characteristics of peers, however, may not be appropriate when social interactions among peers occurred in the previous period and the lag is in fact chosen arbitrarily. Then, the model is not properly identified (Fletcher, 2009).

Another stream of the literature addresses the reflection problem using an instrumental variable approach (e.g., Powell et al., 2005; Gaviar et al., 2001; Fletcher, 2009). Finding a credible instrument predicting  $\overline{peer}_{-i(g)}$  but excluded from the model (1) is difficult if researchers do not provide additional controls for self-selection. Authors typically assume that some of peers' characteristics do not directly affect an individual's decision to take up smoking and use these characteristics as instruments, making the assumption that the contextual effect does not exist ( $\alpha_3$  is equal to zero). The credibility of these excluded instruments can be, of course, questioned. Finally Kremer and Levy (2008) rely on a experiment that randomly assigns college students to their roommates to deal with identification problems.

In this paper, I use an approach similar to that of Kremer and Levy (2008). Although the Czech secondary schooling system does not provide a randomized experiment, it is realistic to assume that the majority of classmates cannot have influenced each other before their enrollment into a secondary school (details about the institutional setting are extensively explained in the next section) Apart from the reflection problem, it is also necessary to address the self-selection into schools and the potential presence of a school specific anti-smoking sentiment.

Self-selection problem arises when a peer group is created based on some common unobserved factors affecting the peers' smoking and the individual's decision to smoke. For example, children with a similar family background that affects their propensity to become a smoker can sort themselves into specific schools. In the context of equation (1) the selection issue is reflected in the correlation of the common part of the error term  $\mathcal{E}_{i(g)}$  with smoking prevalence  $\overline{peer}_{-i(g)}$ .

The most credible solution to this problem is a direct randomized assignment of individuals into peer groups. Randomized experiments are, however, rarely available to study secondary school students as secondary school systems are inherently based on sorting of students into schools based in part on unobservable characteristics. To overcome this selection problem, many recent studies (e.g., Fletcher, 2009 or Lunborg, 2006) use school fixed effects that control for all unobserved characteristics of a school as well as average unobservables of the school's students. Thus, their estimation uses only within-school variation in peers' behavior, which is claimed to be random. This approach can also be questioned, however, because existing evidence shows that students can be non-randomly assigned into classes based on their abilities and other characteristics. If this is the case, the estimates are again biased upward. On the other hand, it is also known that fixed effect estimation in the presence of sorting can cause a downward bias of the estimates due to negative correlation between unobserved and observed characteristics of students within a school (Bayer and Ross, 2006). Thus the total bias of fixed effect analysis is unknown.

I propose an alternative solution, which directly controls for unobserved school-specific characteristics of students using a natural assumption about the choice of school, namely that applicants derive their expectations about the school's smoking attitude based on the smoking prevalence among current enrolled students. The regression analysis thus employs older students' smoking as a proxy for expectations and preferences of fresh students toward smoking. This approach respects the design of the enrollment process and is suitable for the data that are widely available for European countries.

The next section describes the institutional setting of Czech secondary schools and the identification strategy in detail.



### 3. Institutional Setting and Identification Strategy

The Czech secondary school system is characterized by early tracking of students (Brunello and Checchi, 2008). Individuals usually attend their neighborhood elementary school and the majority of Czech youth are enrolled into secondary schools based on their choice and an admission exam administered at the age of 15, after completion of the 9th grade at elementary school.

Secondary schools can be divided into three basic types: academic, vocational, and apprenticeship. Academic and vocational schools usually provide four-year secondary programs<sup>2</sup> and students take a school-leaving exam (the so-called ‘Maturita’) at the end of these programs. The ‘Maturita’ is a prerequisite for tertiary education and obligatory for all students at vocational and academic schools (Jurajda, 2005). Apprenticeship programs do not lead to ‘Maturita’ and apprentices do not apply to colleges and universities, but usually become blue collar workers<sup>3</sup>. As Munich (2004) points out, apprenticeship programs usually draw pupils from the lower end of the ability distribution.

The main difference between academic and vocational schools is in their curriculum. Academic schools provide a general education that prepares graduates for college and university studies. Vocational schools provide an education focused on various fields: technical, business, pedagogical, and healthcare. Their graduates are expected to be ready to enter the labor market as well as colleges in their particular field.

The majority (approximately 80%) of Czech secondary schools are public and do not charge a tuition fee. All secondary schools typically organize their own written entrance

---

<sup>2</sup> Some academic schools also provide an 8 years program, to which students are enrolled after their 5th year at an elementary school. According to the manager of the ESPAD survey in the Czech Republic, it cannot be ruled out that a few classes from the 8 year track schools were included. However, the analysis on the sample not containing the academic schools does not show any main differences in results.

exams, which play a crucial role in the admission process (GPA from primary school is also taken into account). Although information about the admission process is not public, I can employ the following assumptions. Firstly, classes at primary schools are generally heterogeneous in skill distribution and pupils from one class apply to different secondary schools. Secondly, families do not move into new neighborhoods based on the quality of secondary schools. These assumptions are supported by the fact that primary schools are usually not directly linked with any particular secondary school, and the mobility of families is generally fairly low.

The admission process has been recently under reform. My data cover the period 1999 – 2003, when the admission process had the following form. Graduates from primary school send an application to two secondary schools that are of interest to them. These secondary schools then select applicants based on results of entrance exams (and previous GPA). If an applicant is not successful in the first round, he/she enters a second round<sup>4</sup>.

This mechanism has the following implications for the proposed identification strategy. First, there is a low chance that peers from one class at the secondary school could have interacted with each other before they were enrolled into the school. The second implication is that students can choose their schools based on their observed and unobserved characteristics affecting their smoking.

The first implication helps us to solve the reflection problem employing a similar method to Kremer and Levy (2008): by using predetermined smoking instead of current smoking. Thus, the baseline specification has the following form:

---

<sup>3</sup> Apprenticeship programs correspond to the ISCED 2 level, according to the OECD ranking. Secondary schools with 'Maturita' correspond to the ISCED 3A level.

<sup>4</sup> Secondary schools were obliged to leave a certain number of free slots for the second round, and a few schools enrolled students even after the two official rounds were over.

$$(2) \quad \Pr(\text{smoke}_{i(g),t}) = \alpha_0 + \alpha_1 \overline{\text{peer}_{-i(g),t-1}} + \alpha_2 \text{Exp}_{i(g),t-1} + \alpha_3 X_{i(g)} + \alpha_4 \overline{X}_{-i(g)} + \varepsilon_{i(g),t}$$

where  $\overline{\text{peer}_{-i(g),t-1}}$  is the pre-secondary school smoking of peers,  $\Pr(\text{smoke}_{i(g),t})$  is the probability of an individual  $i$  becoming a daily smoker in class  $g$ ,  $\text{Exp}_{i(g),t-1}$  is the past experience of individuals with smoking cigarettes and marijuana, drinking beer and drunkenness; the remaining controls are the same as in the previously discussed model (1) and are time invariant.

The crucial implication of model (2) is the non-existence of reverse causality between pre-secondary school daily smoking among peers<sup>5</sup> and the probability of becoming a daily smoker.<sup>6</sup> In other words, an increase in daily smoking at secondary school cannot cause previous experience of peers' smoking.

Next, the selection problem needs to be addressed. First, specification (2) already controls for individual pre-secondary school behavioral characteristics (experience with a first cigarette, beer, marijuana and drunkenness), which should diminish selection bias based on pre-secondary school experience with risky behavior. The data also allow us to control for family characteristics that do not change over time (education of parents, completeness of family, and the smoking of older siblings). The survey does not sufficiently cover smoking and other risky behavior of parents; I assume that the parental effect is constant over time.

---

<sup>5</sup> The peers' pre-secondary school daily smoking most likely contains a recall measurement error that can bias the results toward zero. This problem cannot be solved in this paper. A potential solution lies in undertaking a longitudinal survey that tracks students over time.

<sup>6</sup> One can also consider a second option to avoid the reflection problem: the instrumental variable approach. Current daily smoking of peers can be instrumented by the pre-secondary school smoking of peers following the suggestion in Powell et al. (2005). The instrumental variable approach is performed as a robustness check and presented in section 5.

However, self-selection can still bias the results if students sort into schools and classes based on their specific unobserved factors. That is, students may choose the secondary school where peers are similar in some unobserved characteristics that are correlated with the potential start of daily smoking. Then the correlation between the probability of becoming a daily smoker and peers' smoking is spurious. Students just indirectly express their preferences toward smoking by their choice of school.

To overcome this problem, I assume that sorting into secondary schools is time invariant. This assumption is common in the current literature and usually results in a fixed-effect analysis (e.g., Lundborg, 2006). Here, however, I modify the fixed effect approach to the available data, which contain not only one class of first-year students, but also one class of third-year students in each school. Importantly, this approach is also directly derived from the decision-making process of applicants about their preferred schools. In particular, the smoking behavior of third-year students is used as a control for the first year students' expectation about their future classmates. The reason is the potentially important role of third-year students' characteristics associated with smoking for the decision of applicants about their secondary schools. To clarify this approach, I consider the following model of smoking decisions and school choices.

In the first stage, individuals gain experience with smoking and related activities (alcohol and marijuana). These experiences are directly included in specification (2) using the vector of pre-secondary school characteristics  $Exp_{i(g),t-1}$ . In the second stage, students choose a secondary school and go through the admission process. The final stage takes place at the secondary school – the decision to become a daily smoker can depend on actually revealed peers' smoking.

The second stage is crucial for the effects of selection bias. The choice of secondary school can depend on various factors: individual preferences for schools, regional supply of secondary schools, quality of secondary schools, individual budget constraints, admission process, etc. The self-selection causes a bias to the extent that applicants choose a secondary school based on their preferences toward smoking. This can also be expressed as a minimization of the difference (Akerlof, 1997) between individual characteristics related to smoking and expected characteristics of future peers. Applicants might choose a school  $S$  that satisfies the following condition:

$$\min \{ smoke_i - E_i[peers_{-i(g),S}] \}$$

where  $smoke_i$  is a probability measure characterizing the propensity of an applicant  $i$  to smoke. It includes all observed and unobserved characteristics related to the current and potential future smoking (e.g., attitude toward smoking).  $E_i[peers_{-i(g),S}]$  is an individual expectation about future peers' characteristics associated with their smoking.

Therefore, if a student has unobserved positive preferences toward smoking (and is likely to become a daily smoker), he/she would prefer to enroll in a secondary school with peers who have similar characteristics associated with smoking, holding all other school characteristics constant. This implies that individuals who choose a particular secondary school have similar expectations about future peers which are driven by their current smoking and by common unobserved characteristics related to initiation of smoking in the future. The individuals' expectations about future peers  $E_i[peers_{-i(g),S}]$  are unobserved, but a possible

source of expectations about future peers could be the behavior of older students at the particular secondary school.<sup>7</sup>

However, the final composition of a class is also influenced by many other factors. In particular, the entrance exams and school policy of assigning students to particular classrooms is out of the control of applicants. Thus, the final composition of peers in a class  $g$  (subset of a school  $S$ ) has the following form:

$$(3) \quad \overline{peer}_{-i(g),S} = \overline{E_i(peer_{-i(g),S})} + \mu_{g,S}$$

where  $\mu_{g,S}$  is an unexpected “prediction” of shock that affects the composition of classroom  $g$  and  $\overline{E_i[peer_{-i(g),S}]}$  is the mean over all students’ expectations, which are formed based on the older students’ characteristics.

Expectations about future peers are approximated by older schoolmates’ lagged smoking, which is an easily available source of information about future peers. The unexpected shock  $\mu_g$  remains in the variation of peers’ smoking to allow for bias-free estimation of  $\alpha_1$ . The estimated coefficient  $\alpha_2$  corresponds to the effect of past individual expectations about peers on the current smoking decision. This can also be interpreted as a neighborhood or a parental effect, because individuals were influenced by them before enrollment into the school. The final model is this:

---

<sup>7</sup> It might be costly or not possible for applicants to search among current peers and at the same time make a correct guess about their future classmates, because they may have different preferences for schools. On the other hand, it is much more efficient to search among already enrolled older students and obtain information about their types.

$$(4) \quad \Pr(\text{smoke}_{i(g),S,t}) = \alpha_0 + \alpha_1 \text{peer}_{-i(g),S,t-1} + \alpha_2 \text{Old}_{S,t-1} + \alpha_3 \text{Exp}_{i(g),S,t-1} + \alpha_4 X_{i(g),S,t} + \alpha_5 \bar{X}_{-i(g),S,t} + \varepsilon_{i(g),S,t}$$

where  $\text{Old}_{S,t-1}$  is older schoolmates' experience with daily smoking (approximated from time  $t-1$  when applicants made their enrollment decision) at secondary school  $S$ . The lagged experience with smoking helps not to confound self-selection with current social interactions between older and younger students. The main advantage of this approach is the ability to directly control for unobserved preferences of students in the secondary schools using available information about older schoolmates.

#### 4. Data Description and Overview of Risky Behavior

The data come from the European School Survey of Alcohol and Other Drugs (ESPAD). This survey primarily consists of 16 year old high school students from 26 European countries who were asked about their tobacco, alcohol, and drug consumption. The survey was performed in four waves: 1995, 1999, 2003 and 2007. The database includes information about current smoking, past smoking and start of daily smoking, consumption of alcohol and marijuana, education of parents, the existence of siblings, use of spare time and the type of school that the student attends, perceived riskiness of smoking, average GPA, measure of self-esteem, and number of family members.

For the purpose of the estimation, I pool data from 1999 and 2003. The sample from 1995 does not contain information about third-year students, which is crucial for my identification strategy, and data from 2007 are not yet available. The sample from 1995 also omits some important characteristics (for example, siblings' smoking) and the average age is different: 15.8 as compared to 16.2 in 1999 and 2003. For these reasons the data from 1995

are not used for this analysis of peer effects.<sup>8</sup> The main quantitative description of smoking behavior is summarized in tables 1-5 (in appendix). The general prevalence of smoking is quite high. Forty-four percent of the sample report having at least one cigarette during the previous 30 days. Daily smoking is reported by 30% of the sample.

The statistics show a high variation of smoking outcomes across types of schools, but low variation across year of collection and gender. For example, the level of females smoking at academic schools is about 30% of that in apprenticeship schools (tables 1 and 2). This suggests that a different selection mechanism and/or social interactions can exist across types of schools. This is reflected in the estimation by controlling for school-types fixed effects.

Generally, a high prevalence of smoking is accompanied by a high consumption of alcohol: 72% of females and 88% of males consumed beer in the last 30 days. The smoking of marijuana during the last 30 days reached 22% of males and 18% of females in 2003. The consumption of beer and marijuana in my specification is used to control for individual pre-secondary school behavior and preferences toward risky behavior.

Table 3 shows the self-reported start of a daily smoking habit. If a respondent reports having started daily smoking younger than 15, it is assumed to be predetermined smoking that occurred before enrollment into the secondary school. Daily smoking initiation that is reported to have begun at the age of 15 and 16 most likely happens at the time of secondary education. Based on this information the key explanatory variable is created: peers' pre-

---

<sup>8</sup> A comparison of 1995 and 1999 statistics reveals high growth of smoking prevalence in all three types of secondary schools. The size of the increase is substantial even after controlling for age, which on average increased between 1995 and 1999 by 0.4 years. Female smoking increased more than male smoking and the level of smoking considerably differs across schools. The proportion of smokers stays approximately the same in 1999 and 2003, the proportion of daily smokers slightly decreased, but still remains high. I leave this phenomenon for future research.



secondary school daily smoking. This variable is summarized in table 4 and displays how different types of schools draw students with different smoking histories.

Table 5 displays the descriptive statistics of third-year students' past experience with daily smoking. This is the key variable controlling for the selection of first-year students into the schools. As described earlier, past daily smoking of older students may correspond to the formation of the prospective students' expectations. A similar variation to first-year students' smoking can be observed across types of schools and years.

The descriptive statistics of all other variables and characteristics of first-year students are in table 6. The structure of samples from 1999 and 2003 is fairly similar in many aspects, including the number of observations and classes, means of all predetermined individual characteristics (for example, age and education of parents), numbers of students in all three types of secondary schools, and geographical structure. For the purpose of estimation I pool these two samples. The time fixed effects used in estimation capture all unobserved differences between first year students surveyed in 1999 and 2003.

Finally, two different samples are presented in table 6: with and without pre-secondary school daily smokers. In fact, there are no important differences in terms of regional structure and other non-behavioral characteristics between these two samples.

## 5. Results

The results are presented in tables 7 and 8. They are from linear OLS regressions determining the individual probability of being a current daily smoker in the first year at secondary school, and all standard errors are clustered at the class level. The first presented result is from "naïve" specifications not controlling for the reflection and self-selection problem; subsequently, the reflection and selection biases are accounted for. The sample is

divided into two parts: females and males, and show that social interaction has a different strength for each gender. The results in table 7 employ the full sample including those who report pre-secondary school daily smoking; while table 8 presents results with a restricted sample that does not contain pre-secondary school daily smokers.

The first specification (1) in table 7 does not control for any pre-secondary school behavioral characteristics and the key explanatory variable is the peers' current daily smoking (individual smoking is always excluded from the peer variable). The other control variables are time fixed effects, current GPA, participation in sports on a daily basis<sup>9</sup>, parental education, older siblings' smoking and school-type fixed effects. The effect of peers' daily smoking on individual daily smoking is significant and positive. The coefficient is bigger for males (0.350) than for females (0.273).

Next, I substitute current smoking with pre-secondary school smoking to alleviate the reflection problem. The cost is the measurement error that can potentially bias the results downward as was discussed above. The estimated peer effects decreased to 0.268 for males and 0.170 for females, respectively.

The next step addresses the selection problem. First, pre-secondary school behavioral characteristics (experience with smoking, marijuana and alcohol) are included in the regression. These controls should also capture the family effect that is directly controlled for only by the education of parents, completeness of family and the smoking of an older sibling.

The results presented in the third row of table 7 show an approximately 40% decrease in the estimate of peer effects: 0.153 as opposed to 0.268 in the previous case for males and insignificant results for females. This suggests that sorting of students into secondary schools

---

<sup>9</sup> In this approach *participation in sports on a daily basis* is not influenced by new peers at secondary school and is predetermined by his/her activity at an elementary school. In the Czech Republic, sport clubs are not

based on pre-school experience with cigarettes, drugs and alcohol strongly biases the peer effects' estimate. As described above, this does not have to capture all the bias, because classmates may have other common characteristics that could influence the individual smoking decision (contextual effect). Therefore, additional controls for other peers' characteristics (with an individual's own level excluded) are included in specification (4): average level of parental education, family completeness, participation in sports on a daily basis, and siblings' smoking. The estimated effect of peers is lower, but the size of the decrease is smaller than in the previous cases. The peer effects for females remain insignificant.

In the final step, the potential sorting of students into schools based on their unobserved attitude toward smoking is considered using the variable *% third year daily smokers*. It is meant to capture those unobserved factors that are related to the choice of school and smoking decision. The pre-secondary school smoking of peers is the key explanatory variable and it can be observed that, by controlling for older students' smoking, the estimated effect of peers slightly decreases (specification (5) in table 7).

The effect of sorting is estimated to be positive and significant, which suggests that the sorting of students into secondary schools can exist. Intuitively, one can also interpret the impact of older students' past smoking as the influence of other school and neighborhood factors that might affect an individual's decision. The estimated effect of peers, however, does not change significantly. It decreases from 0.126 to 0.114 and standard errors are approximately the same.

A robustness check is done using fixed effect analysis. The caveat of the fixed effect analysis is in the limitation imposed by the data. The ESPAD contains only two classes in

different grades (first and third year) for each secondary school. Therefore, I cannot use grade fixed effects, which might be important, because first and third-year students can differ in their unobserved characteristics. The result is found in row 6 of table 7. The estimated peer effects are similar to those that use only first year students. In the last row, I present the result from the sample that excludes all students from academic schools. The rationale for this is that some classes in the sample could be drawn from an eight-year track, which would invalidate the identification strategy dealing with the reflection and selection problem. The size of the estimated peer effects is similar to specification 5, and the standard errors increased from .05 to .06.

Table 8 presents results that use the restricted sample only with students who were not daily smokers before secondary school. By excluding all pre-secondary school daily smokers the self-selection is diminished, because smokers most likely sort into specific schools. The key explanatory variable is again the proportion of pre-secondary school smokers in classes. The analysis therefore estimates how pre-secondary school daily smokers influence those who did not smoke daily before their enrollment. The results presented in table 8 are similar to those estimated by the previous approach presented in table 7.

Another robustness check is the instrumental variable approach that uses the pre-secondary school smoking of peers as an instrument for current peers' smoking (tables 9 and 10). This approach should diminish the reflection problem, similarly to Powell et al. (2005). The first-stage regression suggests that this instrument has a very strong predictive power and is significant at the 1% level. The results for female and male students are the same as those in tables 7 and 8. The estimate of male peer effects together with standard error increases after

---

optionally stay there longer during their high school.

applying the IV approach; a possible explanation may lie in measurement errors in the pre-secondary school smoking variable, or in the endogeneity of the instrument.

The hypothesis often tested in the literature is whether peers' smoking has a non-linear impact on an individual's decision. In order to test it, I use a similar methodology to that of Clark and Loheac (2007) and create dummy variables for each quartile of peers' pre-secondary daily smoking. The results are presented in table 11. The estimates are not significantly different from zero or from each other. Thus the hypothesis that peer effects are linear cannot be rejected.

The next hypothesis I test whether those students who report trying marijuana before being enrolled in secondary school are more likely to be affected by peers' smoking or not. For that purpose, I create a new variable - the interaction of previous experience with marijuana and peers' pre-secondary daily smoking. Although the estimate is positive and relatively high (0.15) for both females and males, it is not statistically significant from zero at the 10% level.

Comparing the results with the current literature is difficult due to institutional setting, but I can claim that the results are rather on the lower end for estimated peer effects. For example, Lundborg (2006) estimated that increasing the number of peers' smoking by 25% increases the probability of smoking by 12 percentage points. Similar results to Lundborg's are presented in Powell et al. (2005) and Fletcher (2009). The magnitude of peer effects estimated in this paper is similar to the one in Clark and Loheac (2007), who also use lagged peers' smoking, but without any appropriate experiment that would assign students into new peer groups. They estimate that the impact of an increase in peers' smoking by 25% on individual smoking is 2.2 percentage points, while the result for Czech male youth is approximately 3 percentage points.

## 6. Conclusion and Policy Implications

In this paper, I use a social interaction framework to determine whether smoking decisions are influenced by the daily smoking of classmates. Several estimation issues are addressed including the endogeneity of school choice, which might be related to the smoking decision.

The main results suggest that smoking decisions are affected by peers' smoking. There are significant endogenous peer effects mainly for male students. This finding has several important implications. Firstly, the decision and enrollment process into secondary schools has not just human capital consequences, but also important health implications. Secondly, public policies that attempt to influence youth smoking in Czech secondary schools can rely on the existence of a social multiplier for male students.

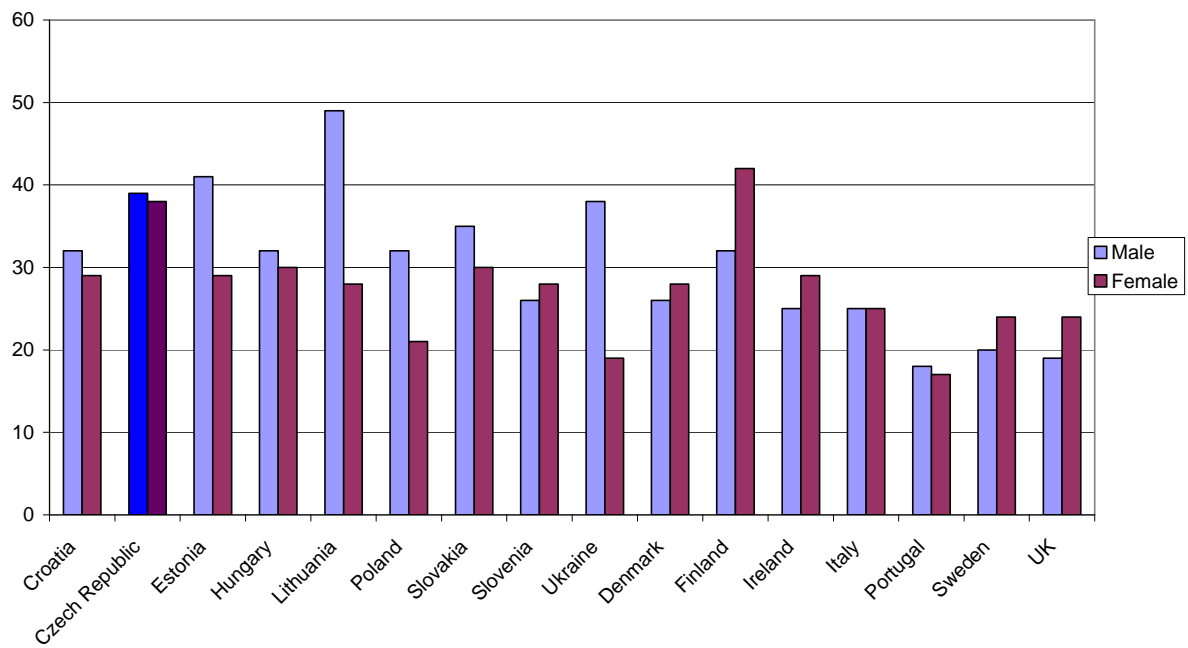
This analysis also has certain limitations. First, the peer group is arbitrarily defined as a class, which may be too narrow. For example, female students might spend time with mates not enrolled in the school. A social multiplier for females thus might exist, but not within a class. Second, the analysis omits several characteristics that might play an important role, namely an individual budget constraint. Although these variables would improve the analysis, they would most likely not change the estimated difference between males and females.

## References

- Akerlof, G.A. (1997): "Social Distance and Social Decision", *Econometrica*, Vol. 65, No. 5, 1005-1027
- Bayer, P. and Ross, A. L. (2006): "Identifying Individual and Group Effects in the Presence of Sorting: A Neighborhood Effects Application", Working paper 12211, National Bureau of Economic Analysis, Cambridge
- Brunello, G. and Checchi, D. (2007): "Does School Tracking Affect Equality of Opportunity? New International Evidence", *Economic Policy*, Vol. 23, 601-649
- Chaloupka, F. and Watner K. (2000): "Economics of Smoking", in *Handbook of Health Economics*, Cuyler A. and Newhouse, J. (Eds.), North-Holland, Elsevier
- Clark, A. and Lohaec, Y. (2007): "'It Wasn't Me, It Was Them!' Social Influence in Risky Behavior by Adolescents", *Journal of Health Economics*, Vol. 26(4), 763-784
- Fletcher, J. M. (2009): "Social Interaction and Smoking: Evidence using Multiple Student Cohorts, Instrumental Variables, and School Fixed Effects", *Health Economics*, in press
- Gaviar, A. and Raphael, S. (2001): "School-Based Peer Effects and Juvenile Behavior" *Review of Economics and Statistics*, Vol. 83, No. 2, 257-268
- Glynn, T.J. (1981): "From Family to Peer: A Review of Transitions of Influence among Drug-using Youth" *Journal of Youth and Adolescence*, No. 10, 365-383
- Jurajda, S. (2005): "Czech Returns to Schooling: Does the Short Supply of College Education Bite?", *Czech Journal of Economics and Finance (Finance a úvěr)*, 55 (1-2), 83-95
- Kremer, M. and Levy, D. (2008): "Peer Effects and Alcohol Use among College Students", *Journal of Economic Perspectives*, Vol. 22, No. 3, 189-206
- Lundborg, P. (2006): "Having the Wrong Friends? Peer Effects in Adolescent Substance Use", *Journal of Health Economics*, Vol. 25, 214-233
- Manski, Ch. (1995): *Identification Problems in Social Sciences*, Harvard University Press: Cambridge, MA
- Münich, D. (2004): "Estimating the Impact of School Quality, Selection and Supply on Students' Achievement: Evidence from the Czech Nation-Wide Testing of Youth", mimeo
- Powell, L., Taurus, J. and Ross, H. (2005): "Importance of Peer Effects, Cigarette Prices, and Tobacco Control Policies for Youth Smoking Behavior", *Journal of Health Economics*, Vol. 24, 950-968
- Sacerdote, B. (2001): "Peer Effects with Random Assignment; Results from Dartmouth Roommates", *Quarterly Journal of Economics*, 116, No. 2, 681-704

## Appendix

Graph 1: Lifetime Use of Cigarettes in 2003 (40 and more), Age 16



Source: European School Survey of Alcohol and Other Drugs, 2003



Table 1: Trying a cigarette within the last 30 days (share, st.dev.)

School Type	1999		2003	
	Male	Female	Male	Female
Academic	.38 (.49)	.32 (.47)	.22 (.41)	.25 (.43)
Vocational	.40 (.49)	.43 (.49)	.42 (.49)	.42 (.49)
Apprenticeship	.55 (.50)	.56 (.49)	.58 (.49)	.69 (.46)
Total	.45 (.50)	.44 (.50)	.44 (.49)	.44 (.49)

Table 2: Current daily smoking (share, st.dev)

School Type	1999		2003	
	Male	Female	Male	Female
Academic	.21 (.41)	.19 (.39)	.08 (.27)	.09 (.28)
Vocational	.28 (.45)	.27 (.45)	.23 (.42)	.20 (.40)
Apprenticeship	.43 (.49)	.45 (.49)	.38 (.48)	.46 (.49)
Total	.32 (.27)	.30 (.46)	.27 (.44)	.24 (.42)

Table 3: Start of daily smoking, first year students (in %)

	1999		2003	
	Male	Female	Male	Female
Never	61.5	61.15	62.9	65.3
11 and earlier	2.8	0.8	2.8	1.0
12	3.3	2.8	3.5	2.8
13	5.1	5.2	5.5	5.3
14	10.2	9.3	9.2	9.1
15	11.6	12.7	11.4	12.4
16 and later	5.5	6.4	4.6	4.2
avg. age	16.2	16.2	16.1	16.1

Note: Self-reported start of daily smoking, individuals are enrolled in secondary school approximately at age 15.

Table 4: Peers' daily smoking before enrollment into secondary school, share (st.dev)

School Type	1999		2003	
	Male	Female	Male	Female
Academic	.15 (.11)	.13 (.10)	.09 (.07)	.09 (.06)
Vocational	.19 (.10)	.15 (.09)	.22 (.10)	.19 (.09)
Apprenticeship	.25 (.13)	.24 (.12)	.32 (.12)	.35 (.12)
Total	.20 (.12)	.17 (.11)	.23 (.14)	.20 (.13)

Table 5: Smoking experience of third year students, share (st.dev)

School Type	1999		2003	
	Male	Female	Male	Female
Academic	.11 (.32)	.14 (.35)	.12 (.33)	.13 (.34)
Vocational	.24 (.42)	.23 (.42)	.22 (.41)	.26 (.44)
Apprenticeship	.33 (.47)	.32 (.46)	.33 (.47)	.37 (.48)
Total	.24 (.43)	.23 (.42)	.25 (.43)	.25 (.43)

Table 6: Descriptive statistics

Variables	1999		2003	
	(1) Mean (st.dev.)	(2) Mean (st.dev.)	(3) Mean (st.dev.)	(4) Mean (st.dev.)
Daily smoker	.31 (.46)	.22 (.42)	.26 (.44)	.16 (.37)
Daily smoker before sec. sch.	.19 (.39)	n.a.	.22 (.41)	n.a.
Try cig. before sec. sch.	.63 (.48)	.55 (.50)	.66 (.47)	.57 (.50)
Drunk before sec. sch.	.33 (.47)	.25 (.43)	.37 (.48)	.27 (.45)
Try marijuana bef. sec. sch.	.26 (.44)	.19 (.39)	.35 (.48)	.26 (.44)
Drink beer before sec. sch.	.70 (.46)	.66 (.47)	.75 (.43)	.70 (.46)
Complete family (1- yes)	.78 (.41)	.79 (.40)	.76 (.42)	.79 (.41)
Sport on daily basis	.29 (.46)	.30 (.46)	.29 (.46)	.31 (.46)
Father's college degree	.21 (.41)	.21 (.41)	.24 (.43)	.26 (.44)
Father's hs degree	.26 (.44)	.27 (.44)	.27 (.44)	.27 (.44)
Age	16.21 (.40)	16.22 (.40)	16.19 (.41)	16.19 (.41)
Female	.52 (.50)	.53 (.50)	.53 (.50)	.53 (.50)
GPA12	.43 (.50)	.47 (.50)	.40 (.49)	.44 (.50)
GPA34	.46 (.50)	.44 (.50)	.45 (.50)	.43 (.49)
GPA56	.09 (.29)	.08 (.27)	.11 (.31)	.08 (.28)
Older siblings smoker	.31 (.46)	.32 (.47)	.30 (.46)	.32 (.47)
Vocational school	.45 (.50)	.46 (.50)	.40 (.49)	.41 (.49)
Academic school	.22 (.42)	.24 (.43)	.26 (.44)	.30 (.46)
Apprenticeship	.33 (.47)	.30 (.46)	.34 (.48)	.29 (.46)
Regions:				
Prague	.11 (.31)	.10 (.30)	.10 (.30)	.10 (.30)
Central	.11 (.31)	.11 (.31)	.10 (.30)	.10 (.30)
South	.07 (.25)	.07 (.25)	.06 (.25)	.06 (.24)
West	.07 (.26)	.07 (.26)	.08 (.28)	.08 (.27)
North	.13 (.33)	.12 (.33)	.13 (.34)	.13 (.34)
East	.13 (.33)	.13 (.34)	.15 (.36)	.15 (.35)
Southeast	.19 (.40)	.19 (.39)	.17 (.38)	.17 (.38)
Northeast	.20 (.40)	.20 (.40)	.20 (.40)	.20 (.40)
Total number of observations	4676	3787	4622	3612
Number of classes	224	224	180	180

Note: samples (1) and (3) are full, samples (2) and (4) do not involve pre-secondary school daily smokers

Table 7: Estimation of peer effects (full sample)

	Peers' smoking	Controls for selection	Female	Male
(1)	% current daily smokers		.273*** (.052)	.350*** (.049)
(2)	% pre-school daily smokers		.170*** (.056)	.268*** (.054)
(3)	% pre-school daily smokers	Individual pre-school behavior	.019 (.050)	.153*** (.053)
(4)	% pre-school daily smokers	(3) + peers' characteristics	.016 (.50)	.126** (.046)
(5)	% pre-school daily smokers	(4) + % older students' lagged smoking	.024 (.054)	.114** (.047)
(6)	% pre-school daily smokers	(4) + school fixed effect	.062 (.045)	.142*** (.043)
(7)	% pre-school daily smokers	(5) no academic sch.	.032 (.70)	.117* (.061)

Note: Results come from OLS regressions. All specifications control for GPA, parental education, family completeness, school type, time and regional dummies. Standard errors are clustered on class level (in fixed effect estimation on school level)

Table 8: Estimation of peer effects (no pre-secondary school daily smokers)

	Peers' smoking	Controls for selection	Female	Male
(1)	% current daily smokers		0.242*** (0.053)	0.310*** (0.053)
(2)	% pre-school daily smokers		0.128* (0.069)	0.227*** (0.066)
(3)	% pre-school daily smokers	Individual pre-school behavior	0.063 (0.066)	0.198*** (0.063)
(4)	% pre-school daily smokers	(3) + peers' characteristics	0.050 (0.067)	0.173*** (0.061)
(5)	% pre-school daily smokers	(4) + % older students' lagged smoking	0.062 (0.067)	0.157** (0.061)
(6)	% pre-school daily smokers	School fixed effect	0.024 (0.066)	0.114* (0.0675)
(7)	% pre-school daily smokers	(5) no academic sch.	.056 (0.782)	.146** (.071)

Note: Results are from linear probability OLS regressions. All specifications control for GPA, parental education, family completeness, school type, time and regional dummies. Standard errors are clustered on class level (in fixed effect estimation on school level)

Table 9: Instrumental variable estimation

	Naive (female) (3)	IV (female) (4)	Naive (male) (5)	IV (male) (6)
% current daily smokers	0.151*** (0.048)	0.060 (0.094)	0.218*** (0.053)	0.281*** (0.074)
% older students lagged smokers	0.069 (0.043)	-0.026 (0.046)	0.112*** (0.034)	0.102*** (0.032)
Individual pre-school behavior	X	X	X	X
GPA and Parental Education	X	X	X	X
School type, time and regional fixed effects	X	X	X	X
Peers' characteristics	X	X	X	X
Observations	4514	4514	4079	4079
R-squared	0.35	0.35	0.23	0.23

Note: The instrument is peers' pre-secondary daily smoking. The Instrumented variable is % *current daily smokers*. I control for selection using various pre-secondary school individual behavioral characteristics, current peers' characteristics and older schoolmates lagged behavior (as a proxy for expectations).

Table 10: Instrumental variable – first stage

	Female (2)	Male (3)
% pre-school daily smokers	0.571*** (0.015)	0.558*** (0.016)
Individual pre-school behavior	X	X
GPA and Parental Education	X	X
School type, time, reg. fixed effects	X	X
Peers' characteristics	X	X
Observations	4515	4081
R-squared	0.62	0.48

Note: Explained variable is the current share of daily smokers.

Table 11: Testing for non-linearity of peer effects

	Male	Female
Peers' smoking:		
2. quartile	0.017 (0.017)	0.004 (0.022)
3. quartile	0.006 (0.018)	0.033 (0.023)
4. quartile	0.020 (0.023)	0.033 (0.026)
Individual pre-school behavior	X	X
GPA and Parental Education	X	X
School type, time, reg. fixed effects	X	X
Peers' characteristics	X	X
Observations	3648	3206
R-squared	0.24	0.17

Table 12: Testing for peer effects for smokers of marijuana

	Female	Male
Try marijuana* % pre-school daily smokers	0.150 (0.137)	0.153 (0.121)
Try marijuana before sec. school	0.271*** (0.038)	0.208*** (0.032)
% pre-school daily smokers	0.014 (0.059)	0.10* (0.058)
Individual pre-school behavior	X	X
GPA and Parental Education	X	X
School type, time, reg. fixed effects	X	X
Peers' characteristics	X	X
Observations	3644	3195
R-squared	0.24	0.17