Melbourne Institute Working Paper Series

Working Paper No. 27/12

Peer Effects in Adolescent Cannabis Use:
It's the Friends, Stupid

John Moriarty, Duncan McVicar and Kathryn Higgins

MELBOURNE INSTITUTE®
of Applied Economic and Social Research

Peer Effects in Adolescent Cannabis Use: It's the Friends, Stupid*

John Moriarty[†], Duncan McVicar[‡] and Kathryn Higgins[§]

[†] Institute of Child Care Research, Queen's University Belfast
[‡] Melbourne Institute of Applied Economic and Social Research,
The University of Melbourne
[§] School of Sociology, Social Policy, and Social Work, Queen's University Belfast

Abstract

This paper examines peer effects in adolescent cannabis use from several different reference groups, exploiting survey data that have many desirable properties and have not previously been used for this purpose. Treating the school grade as the reference group, and using both neighbourhood fixed effects and IV for identification, we find evidence of large, positive, and statistically significant peer effects. Treating nominated friends as the reference group, and using both school fixed effects and IV for identification, we again find evidence of large, positive, and generally statistically significant peer effects. Our preferred IV approach exploits information about friends of friends – 'friends once removed', who are not themselves friends – to instrument for friends' cannabis use. Finally, we examine whether the cannabis use of schoolmates who are not nominated as friends – 'non-friends' – influences own cannabis use. Once again using neighbourhood fixed effects and IV for identification, the evidence suggests zero impact. In our data, schoolmates who are not also friends have no influence on adolescent cannabis use.

1. Introduction

Early adolescent cannabis use has been linked to a variety of negative consequences for the user including poorer educational outcomes (e.g. van Ours and Williams, 2009), while persistent use in adolescence has been shown to increase the probability of adult onset psychological ill-health including psychotic symptoms (e.g. Patton et al., 2002; Kupper et al., 2011). Parents and policy makers may therefore want to discourage adolescents from using cannabis. Evidence on the factors that influence adolescent cannabis use can help in this regard, and although some factors are unlikely to be manipulable (e.g. gender, household structure), others may be more open to intervention from policy makers or parents (e.g. price, choice of school or friends).

This paper provides new evidence on observable factors that are associated with adolescent cannabis use. In particular, we examine evidence for peer effects in adolescent cannabis use using school-based survey data from Northern Ireland not previously used for this purpose. It is widely believed that peer behaviour is one of the key factors influencing whether an adolescent uses cannabis or other substances. But such peer effects – the causal links between peer behaviour and individual behaviour that Manski (1993, 2000) calls *endogenous social interactions* – are notoriously hard to quantify. Despite a growing literature that seeks to estimate peer effects in cannabis use and other adolescent substance use behaviours, the evidence remains inconclusive, at least in terms of magnitude, even if most studies find evidence of non-zero peer effects.

Another issue that studies of peer effects in adolescent substance use must confront is the question of how to define a "peer" and which of the available reference groups is the most relevant in determining individual behaviour. Partly because of data availability, many studies have used school-based survey data, where the school, school grade or school class is treated as the reference group (e.g. Gaviria and Raphael, 2001; Powell et al., 2005; Lundborg, 2006; Fletcher, 2010; McVicar, 2011). For behaviour that mostly takes place outside of school such as cannabis consumption, however, these may not be the only (or even the most) relevant reference groups. In particular, one can imagine that an adolescent's *friends* will act as an important reference group in decisions about substance use. To date, however, few studies have been able to provide credible estimates of peer effects in adolescent substance use between friends. The main problem is that suitable data are scarce, with a number of studies forced to rely on individuals' *perceptions* about friends' behaviours (e.g. Kawaguchi,

2004; Krauth, 2005; Krauth, 2007; Vasquez, 2010). Such an approach introduces additional identification issues into the mix, however, including the likelihood that an individual will project his or her own behaviour onto peers when reporting perceived peer behaviour (Norton et al. 2003; Lundbrog, 2006; Vasquez, 2010).

There is also a question over whether estimates of peer effects in substance use from schoolmates reflect the effects of *all* schoolmates or some subset of schoolmates, i.e. *friends*. McVicar and Polanski (2012) argue that because at least some of an adolescent's friends are likely to be in the school, in models where schoolmates are assumed to be the sole reference group, schoolmate behaviour may be proxying, in part or in full, for friends' behaviour. This is a critical distinction for parents concerned with minimising exposure to negative peer influences. If schoolmates in general impact on behaviour, then by exercising school choice parents may reduce such negative influences. If it is only friends that impact on own substance use behaviour, however, then efforts to influence the adolescent's selection of friends within the school may have a bigger impact on substance use behaviour than school choice. Even if *both* friends *and* other schoolmates impact on own behaviour, the relative magnitude of these influences is potentially important information for parents. This issue can only be resolved by estimating models that allow for peer effects from friends and other schoolmates simultaneously, and although McVicar and Polanski (2012) make some progress in this regard, they are constrained by only having perceived data on friends' substance use in addition to data on classmates' use.

In this paper we exploit the desirable properties of our survey data to make significant contributions to the substance use peer effects literature in the following three respects. First, treating the school grade as the reference group, we use a combination of neighbourhood fixed effects and instrumental variables (IV) to provide estimates of peer effects in adolescent cannabis use that are arguably better identified than many of those in the literature. The argument hinges on a reasonable ex ante case for instrument excludability, which is further strengthened by the inclusion of neighbourhood fixed effects which plausibly reduce validity problems associated with endogenous sorting into schools.[1] The resulting estimates suggest large, positive, and highly statistically significant peer effects in adolescent cannabis use between schoolmates.

---

[1] For more detail on this argument see Lundborg (2006), Fletcher (2010) and McVicar (2012).

Second, we exploit information on nominated friends within the school grade to estimate peer effects between friends that do not rely on perceived friends' behaviour. Very few studies have been able to do this previously. One exception is Norton et al. (2003), who examine peer effects in adolescent tobacco smoking using information on up to three nominated 'best friends' contained in survey data for North Carolina. Given that they provide only OLS estimates, however, the identification problems set out by Manski remain an issue, and we cannot be sure of the extent to which the resulting statistical associations capture causal relationships between reference group and own behaviour. A reference group of no more than three peers also seems rather restrictive. Another exception is Clark and Lohéac (2007), who examine peer effects in a number of adolescent substance use behaviours, including cannabis use, using information on up to ten nominated friends contained in the Add Health study. Despite lagging peer behaviour and including school fixed effects, however, their estimates may still be subject to biases arising from selection of friends. Here we go further than these earlier studies by exploiting additional information on network structure within schools to provide IV estimates of peer effects from friends using information on friends of friends who are not themselves friends. Given that the individual is not directly exposed to the characteristics of friends of friends who are not themselves her friends, we assume information about this group can be used to instrument for friends' behaviour. Again, the resulting estimates suggest large, positive, and highly statistically significant peer effects in adolescent cannabis use between friends.

Third, we use information on friendship nominations to split school grades into two separate reference groups – friends and 'non-friends' – and then simultaneously estimate peer effects from both groups. In doing so we build on McVicar and Polanski (2012) in the crucial respect that behaviour among friends is observed directly rather than based on individuals' reported perceptions of friends' behaviour. We find strong evidence that schoolmates not nominated as friends have no impact on an individual's cannabis use. The suggestion is that estimates of peer effects in adolescent substance use that specify a single school/grade/class-level reference group may simply be picking up peer effects from friends within the school rather than from schoolmates in general. This finding also provides additional support for the 'friends-once-removed' IV strategy used earlier to identify peer effects from nominated friends.

The remainder of this paper is set out as follows. The following section briefly reviews the existing literature providing estimates of peer effects in adolescent substance use generally

and adolescent cannabis use in particular. Section 3 introduces our data, taken from the Belfast Youth Development Study (BYDS). Section 4 sets out our empirical approach and discusses identification issues. Section 5 presents and discusses estimation results. Section 6 concludes.

2. Existing Studies

For conciseness, we concentrate on two sets of studies. First, we consider studies that seek to quantify peer effects in school-age adolescent substance use, and cannabis use in particular, where the school, grade, or class is treated as the reference group. We restrict further to studies that have an explicit strategy for dealing with the identification issues raised by Manski (1993, 2000) and that have been published from 2001 onwards. Second, we consider studies that seek to quantify peer effects in school-age adolescent substance use, and cannabis use in particular, specifically from friends. For this group, we relax the identification restriction so as not to exclude the bulk of the literature.

First consider studies that treat the school, school grade, or school class as the reference group. Studies that provide estimates of peer effects in cannabis use (or a wider definition of illicit drug use that includes cannabis use) include Gaviria and Raphael (2001), Lundborg (2006), Clark and Lohéac (2007), and McVicar and Polanski (2012). Gaviria and Raphael (2001) (drug use, school, US) and McVicar and Polanski (2012) (cannabis use, school class, UK) both use IV methods for identification, estimating peer effects coefficients of .32 and .61, respectively. Lundborg (2006) (drug use, school class, Sweden) uses school fixed effects and school fixed effects in combination with IV for identification, estimating peer effects coefficients of .07 and .17 respectively. Clark and Lohéac (2007) (cannabis use, school grade, US) use lagged peer behaviour together with school fixed effects for identification, estimating peer effects of .06 (boys) and -.03 (girls), both of which are statistically insignificant. All four papers also present peer effects estimates, using similar identification strategies, for other substance use behaviours including tobacco and alcohol use. A number of further studies use similar methods to estimate peer effects specifically for adolescent tobacco smoking, including Powell et al. (2005), Soetevent and Kooreman (2007), Sen (2009), Fletcher (2010), McVicar (2011) and McVicar (2012).

Second, consider studies that treat friends as the relevant reference group, but that are limited to using information on perceived behaviour of friends. Peer effects in cannabis or drug use are estimated in this way by Kawaguchi (2004) (US) and McVicar and Polanski (2012) (UK). Kawaguchi (2004) provides simple probit estimates (.14), school fixed effects estimates (.14), and sibling fixed effects estimates (.03). McVicar and Polanski (2012) provide simple probit estimates (.35) and school fixed effects estimates (.33). Both studies also use the same methods to estimate peer effects for other substance use behaviours. Other studies estimate peer effects in tobacco smoking from perceived friends' behaviours, including Norton et al. (2003 (OLS, US), Krauth (2005) (Canada), and Krauth (2007) (US) (both using a sample selection model)). Note that the kind of IV approach used in the school-based studies discussed above is generally not possible in the case of perceived friends' behaviours because information on friends' background characteristics is typically not collected in the surveys on which these studies are based.

Third, some studies treat friends as the relevant reference group and that are able to exploit information on friendship links within the data set to explicitly identify friends. Eisenberg (2004) and Clark and Lohéac (2007) are the two studies that use these kinds of data (both using US Add Health data) to estimate peer effects in cannabis use (among other substance use behaviours). Eisenberg (2004) exploits the mobility of school friends, either from switching schools or from graduating, to provide quasi-experimental estimates of peer effects from nominated friends within school (.12 and .05 respectively). Clark and Lohéac (2007) again use lagged peer behaviour with school fixed effects for identification, estimating peer effects of .12 (boys) and .11 (girls). The only other in-scope study that exploits nominated friends data to estimate peer effects in substance use is Norton et al. (2003) (tobacco smoking, three best friends within school), but only OLS estimates are reported.[2]

3. The Belfast Youth Development Study

The BYDS is a longitudinal study of youth behaviour that tracks pupils in a single year group across a sample of 42 schools[3] in Northern Ireland from age 11/12 years through to the final

---

[2] The main contribution of Norton et al. (2003) is to set out and demonstrate the additional identification problems that arise when using perceived friends' behaviour data in place of actual friends' behaviour.
[3] Questionnaires were also received from a small high-risk booster sample of individuals in alternative non-school education settings. As these individuals attend multiple providers they do not have a clear school-based reference group and are omitted from the sample here.

year of compulsory secondary education at age 15/16 years (Year 8 to Year 12). The first wave of data was collected in 2001 and the final (in-school) wave was collected in 2005.[4] In this paper we focus on a single cross-section taken from the BYDS – wave 3, aged 13/14, collected in 2003. There are 4,459 young people in the wave 3 sample, and after dropping 42 individuals who do not provide any information on cannabis use, we are left with a sample of 4,417. Questionnaires are completed in school under exam conditions, placed by the students in sealed envelopes and collected by staff. Participation was contingent in the first instance on the agreement of school principals. Pupils who did not wish to participate could leave the questionnaire blank. Parents were informed of the school's participation and sent a consent withdrawal slip. If this was not returned, parents' consent was assumed. Overall, very few pupils or parents refused to participate (5% of the total sample, ranging from 0% to 11% in individual participating schools by wave 3). There were also some absences on the day of the survey (8% of the total sample, ranging from 1% to 21% in individual participating schools). The BYDS data make no claim to being fully representative of the relevant population, but all points on regional scales of affluence and deprivation are well-represented, with an over-representation of the most deprived areas.

Our analysis uses information on an individual's cannabis use provided by the answers to two survey questions. First, we construct a binary dummy for cannabis use – taking the value 1 if the individual reports having used cannabis at least once, and 0 otherwise – using answers to the following question: *"Have you tried cannabis in the last year?"* This binary approach is the most common approach in the literature (e.g. Gaviria and Raphael, 2001), although some studies report on substance use in the last 30 days or month rather than the last 12 months (e.g. Clark and Lohéac, 2007). BYDS does not ask about cannabis use over the shorter time period, but there is a follow-up question on frequency of use, as follows: *"Thinking about cannabis, which of these statements best describes you: I have only used it once; I have used it between 2 and 5 times; I use it about once every month; I use it every week; I used to take it but don't anymore?"* We use this second question to construct a dummy for 'at least monthly use', with the expectation that estimates of peer effects based on this second dummy may be more directly comparable with those in the existing literature based on use in the past 30

---

[4] For further information on the BYDS, and on cannabis use within the BYDS sample, see McCrystal et al. (2007).

days/month. 28% of the sample report having used cannabis at least once in the last 12 months, and 11% report using cannabis at least monthly.[5]

A particularly attractive characteristic of the BYDS data is that each individual is asked to nominate up to 10 friends within his or her school grade. First and foremost, this allows us to examine peer effects where nominated friends, as an alternative to the school grade, are treated as the reference group. In most cases these nominated friends are also sample members – we label these 'valid friends' – who in turn nominate their own friends, so the researcher can construct friendship networks within the school. As discussed in the following section, this network structure information can help us identify peer effects from friends by using the characteristics of 'once-removed friends' – friends of friends who are not themselves my friends – to instrument for friends' behaviour. To the best of our knowledge, the only other large-scale, school-based survey data set that has similar network information is the Add Health study for the US, and this aspect of the Add Health data has been increasingly exploited by researchers in recent years (e.g. Clark and Lohéac, 2007; Calvo-Armengol et al., 2009).

Reference group behaviour is captured by variables denoting the proportion of the reference group that report using cannabis in the last 12 months (and similarly for monthly use). Here, exploiting the network information contained in the BYDS data, we examine three reference groups: the school year group (minus the individual), the individual's nominated friends, and grade-mates who are not nominated as friends (hereafter 'non-friends').[6]

In addition to information on use of cannabis and other substances, the BYDS contains data on a host of individual and family background characteristics, as well as limited information at the school level, which we use to construct control variables for the regression analyses discussed in the following sections. These include controls for gender, age, number of friends nominated, number of brothers and sisters, household structure (single parent, step family, 'alternative family'[7] or two biological parents), whether the individual reports having older

---

[5] This is the first year where cannabis use becomes widespread in the BYDS sample. At age 12/13 years only 15% of the sample report having used cannabis over the previous year.

[6] The average number of reported friends in the sample is 7.35, but some sample members do list 10 friends (the maximum permitted by the questionnaire). Although convenient, 'non-friends' is therefore a partial misnomer because, for those individuals constrained by the maximum number of nominations, some of those not nominated as friends may still be viewed as friends by the individual concerned. Our conjecture, however, is that individuals list their friends roughly in order of the perceived 'closeness' of the friendship.

[7] We define alternative families as families that report household structure information but do not fit into any of the other categories.

school friends (outside of the grade), whether the individual reports having older non-school friends (outside of the school), whether they expect to complete school (i.e. to continue in school until the end of Year 12)[8], dummies for parental employment (mother part-time, mother full-time, similarly for father), and whether the school attended is a single-sex school, a grammar school (academically selective), and a Catholic school. We generate binary dummies for missing values for the subset of variables where they are an issue.[9] Table 1 reports sample means for these controls and the cannabis use variables.

<Table 1 here>

4. Estimation Approach and Identification

4.1. Treating the School Grade as the Reference Group

Initially we treat the school grade (year group) as the reference group, following Clark and Lohéac (2007) and Fletcher (2010). We estimate the familiar linear-in-means model of Gaviria and Raphael (2001), as follows:

$$(1) \qquad\qquad y_{is} = \alpha + \delta \overline{y}_{-is} + X_{is}'\beta + \varepsilon_{is},$$

where y_{is} is a binary dummy equal to 1 if individual i reports using cannabis in the last 12 months and equal to 0 otherwise, $\overline{y}_{-is}$ denotes the cannabis use of schoolmates in the same year group (proportion of grade-mates using cannabis, minus the individual) and X_{is} is a vector of controls. The parameter of interest is δ, which captures the association between own cannabis use and the prevalence of cannabis use in the school grade. Equation (1) makes the assumption that peer characteristics do not directly impact on own behaviour, i.e. there are no *contextual effects*. This assumption allows us to provide initial estimates of δ using OLS and also, later, to use excluded schoolmate characteristics as instruments.[10] Gaviria and Raphael

---

[8] At the time of the survey, young people were legally entitled to leave school when they turned 16 years, which for some would have fallen before the end of Year 12.

[9] These control for non-completed questions or questions with 'not applicable' or 'don't know' options (we treat all these as a single missing category). The most missing values are for the father employment dummies.

[10] If peer characteristics influence individual behaviour, then a linear model cannot separately identify the impact of peer behaviour from that of peer characteristics, at least not without IV. Such contextual effects are

(2001) and Powell et al. (2005) justify this assumption by the argument that social interaction between schoolmates takes place mostly at school, away from potential peer family-background influences.

There are two main identification issues in estimating peer effects using a model such as (1), however, as is now well known following Manski (1993, 2000). First, own cannabis use and peer cannabis use can be correlated for a whole host of reasons other than peer effects, including common unobserved factors (e.g. a drug dealer outside the school gates) and endogenous sorting into schools. Such factors imply that peer behaviour may be endogenous, and OLS estimates are likely to be upwards biased as a result. Second, because I am a peer of my peers, my cannabis use may affect my peers' cannabis use at the same time as my peers' cannabis use may affect my cannabis use, i.e. Manski's reflection problem. This not only introduces a potential simultaneity bias to the OLS estimate of δ, but also makes a linear-in-means model unidentified in the presence of contextual effects (hence the Gaviria and Raphael assumption).[11]

As a first step in trying to address these identification issues, we use information on the postcode district[12] of the home address for each individual to augment X_{is} by constructing neighbourhood dummies to control for neighbourhood-level correlated effects (e.g. presence of neighbourhood drug dealers, unobserved socio-economic characteristics of those that sort themselves into the neighbourhood).[13]

Even with the postcode dummies included, however, we still need a method for dealing with simultaneity bias, any remaining selection issues at the school level within neighbourhoods, and which allows us to relax the strong assumption of no contextual effects. We therefore follow Gaviria and Raphael (2001) and others by using observed information about schoolmates to instrument for their cannabis use (estimating by two stage least squares). Broadly along the lines of Fletcher (2010) and McVicar (2012) we present estimates using

---

possible for at least two reasons: endogenous sorting into schools and direct contextual effects (e.g. from peer background characteristics if social interaction between schoolmates occurs outside of school, or from other characteristics of classmates (e.g. gender, age) that may directly influence an individual's behaviour within school).

[11] If there are contextual effects then δ in (1) is not interpretable as the endogenous peer effect (the impact of peer behaviour on my behaviour) but as a combination of the endogenous peer effect and contextual effects, i.e. peer effects 'writ large' (see Hoxby, 2000).

[12] Postcode districts are the first 3 or 4 digits of a UK postcode. There are around 80 postcode districts in Northern Ireland with an average population of around 20,000 individuals.

[13] Postcode information is only available for four fifths of the sample, however, so we trade-off greater control for reduced sample size when these dummies are included. For comparison purposes we also report estimates for the model without postcode dummies estimated on the reduced sample.

the proportion of grade-mates who report having older non-school friends as a single instrument. We know from a controls-only regression (see Table 2 in the next section) that having older non-school friends is a strong predictor of own cannabis use, so the instrument is likely to be highly correlated with peer behaviour. So, to the extent that this instrument can be treated as exogenous, the resulting IV estimates can give consistent estimates of δ. Note that this IV approach also allows us to relax the assumption of no contextual effects by introducing variables for the grade-level means (again excluding the individual) of all the other observables contained in X_{is} into (1).

Our *ex ante* case for the exogeneity of our proposed instrument – the proportion of schoolmates that report having older non-school friends – is as follows. First, whether a 13/14 year-old adolescent has older non-school friends seems less likely to be correlated with parents' decisions to locate in a particular neighbourhood or to choose a particular school than, say, parental education levels (one of Gaviria and Raphael (2001)'s instruments). Second, the neighbourhood dummies give us an additional level of control for endogenous sorting into neighbourhoods and neighbourhood schools. Third, if we accept the argument of Gaviria and Raphael (2001) and Powell et al. (2005) that schoolmates do not generally come into contact with one another outside of school, then direct contextual effects from schoolmates having older non-school friends can be ruled out. One way of testing the empirical support for this argument is to examine the joint significance of all the other peer-level characteristics variables in the IV version of (1) (see Section 5.1). Finally, the fact that the instrument is uncorrelated with all but one of the observed individual characteristics (alternative family household structure) suggests it is also unlikely to be strongly correlated with *unobserved* individual characteristics.[14]

With cross-section data, and with only one grade per school[15], this is essentially as far as we can get in terms of identification where the school grade is treated as the reference group.

4.2. Treating Friends as the Reference Group

Next we exploit information on the network structure within school grades – specifically the nominated friends of each individual – to explore the extent to which these nominated friends influence own behaviour, along the lines of Clark and Lohéac (2007). Again we specify a

---

[14] There are 15 observed individual-level covariates, so we would expect a significant correlation at 95% for one or these 15 covariates at random.

[15] This prevents us including school fixed effects.

linear-in-means model – given by (2) below – but with reference group behaviour now denoted by $\bar{y}_{-is}^{f}$, where the superscript f denotes the set of nominated friends:

$$(2) \qquad\qquad y_{is} = \alpha + \delta \bar{y}_{-is}^{f} + X_{is}^{'}\beta + \varepsilon_{is}$$

As before, we initially assume no contextual effects, so we can provide initial estimates of δ using OLS. We face the same identification issues as above, but arguably selection bias will be a more acute problem where nominated friends are treated as the reference group instead of schoolmates (e.g. Clark and Lohéac, 2007). And although school fixed effects can help reduce selection bias – school fixed effects are now possible because we have multiple reference groups per school – within-school selection of friends on unobservables may still lead to upwards bias in within-school estimates of peer effects. In one respect the reflection problem may also be exacerbated because reference groups based on nominated friends are smaller than school grades, and so may be more impacted by the individual (i.e. simultaneity bias may be empirically more important). On the other hand, not all friendship nominations are reciprocated, and if those we do not nominate do not impact on our behaviour, then there is no reflection problem for unreciprocated nominations. Calvo-Armengol et al. (2009) argue that the other implication of the reflection problem – the fact that endogenous and exogenous peer effects cannot be separately identified in a linear model without IV – is eluded in such a set-up, even with reciprocated links, because reference groups are individual-specific. In principle it is therefore possible to include contextual effects in an extended version of (2) and still obtain OLS estimates of the endogenous peer effect.

As for the schoolmates model, we also estimate (2) using an IV approach. We first present estimates using the same older non-school friends instrument as in the schoolmates model (although specified for nominated friends rather than all grade-mates). Concerns regarding validity in the light of the potential for endogenous sorting and direct contextual effects, however, are more acute here, although as for the schoolmates case, all other friends' characteristics covariates can be included in (2) as controls for contextual effects and their joint significance examined.[16] A more attractive IV strategy is therefore to further exploit the

---

[16] The Gaviria and Raphael (2001) argument that you don't come into contact with schoolmates' families or other outside-school influences is weaker in the case of friends. The instrument is also correlated with whether the individual reports having older non-school friends, whether we include school fixed effects or not.

network structure in the data by using information on friends' *friends*, who are *not* themselves my friends, to instrument for friends' behaviour. In doing so we can further relax the assumption of no contextual effects: we still assume no contextual effects from those who are not directly linked to the individual in the network (i.e. anyone other than ego's nominated friends), but we can now allow for an additional contextual effect from those that are directly linked to me in the network (my friends) having older non-school friends. To the best of our knowledge, ours is the first study to use this IV approach to estimate peer effects in adolescent substance use.[17]

4.3. Separately Identifying Friends from Other Schoolmates

Our final model splits each school grade into those nominated by the individual as friends and those not nominated, who we call 'non-friends'. We then examine peer effects from each of these two (mutually exclusive) reference groups. The relevant model is given by (3), where the $-f$ superscript denotes non-friends:

$$(3) \qquad y_{is} = \alpha + \delta_1 \bar{y}^{f}_{-is} + \delta_2 \bar{y}^{-f}_{-is} + X'_{is}\beta + \varepsilon_{is}$$

The coefficients of interest are δ_1 and δ_2. As in (2), δ_1 captures the peer effect in cannabis use from friends. But δ_2 captures the peer effect from all those in the school grade *not* nominated by the individual as friends. Our hypothesis, building on the preliminary findings of McVicar and Polanski (2012), is that non-friends' behaviour has no impact on own behaviour, i.e. that δ_2=0.

As before, we initially estimate (3) by OLS. To address the remaining identification issues outlined above, we then take a similar IV approach to estimating (3), again using the proportion of friends that report having older non-school friends to instrument for friends' behaviour, and now also using the proportion of non-friends that report having older non-school friends to instrument for non-friends' behaviour. Note that we cannot use the friends-once-removed IV strategy for (3), because friends once-removed are, by definition, non-

---

[17] In this case, the 'once-removed' instrument is uncorrelated with whether an individual reports having older non-school friends, but is correlated with gender, alternative family, and having older school friends.

friends. As before, we can allow all other friends' characteristics and non-friends' characteristics to enter as contextual effects in the IV version of (3).

5.   Results and Discussion

We start by presenting estimates from a controls-only version of (1), estimated by OLS. Results are presented in Table 2, and they confirm our priors based on findings reported by earlier studies using other data sets. Males are more likely than females to report cannabis use (e.g. Clark and Lohéac, 2007; McVicar and Polanski, 2012). Cannabis use is positively associated with age within this age group (e.g. Lundborg, 2006; McVicar and Polanski, 2012). Coming from a single parent household (e.g. Gaviria and Raphael, 2001; Clark and Lohéac, 2007) or step household is associated with higher propensity to report cannabis use. Parental employment does not appear to be strongly associated with cannabis use (e.g. Clark and Lohéac, 2007). We also have information in the BYDS which allows us to include controls that have not previously been included in the cannabis peer-effects literature. Wanting to leave school at 16 (or not answering this question) is strongly positively associated with reported cannabis use. Having older friends, whether in or out of school, is also positively and strongly associated with reported cannabis use. Finally, we include three school-level controls. Attending a single-sex school is not significantly associated with cannabis use. Attending a grammar school, however, is strongly negatively associated with cannabis use.[18]   Attending a Catholic school is associated with a higher propensity to report cannabis use.[19]

<Table 2 around here>

5.1. Treating the School Grade as the Reference Group

Next consider estimates of peer effects obtained by including school-grade cannabis use in (1) alongside the controls discussed above. Results are presented in Table 3. Estimating (1) by OLS suggests a strong, highly significant and positive association between own cannabis

---

[18] Grammar schools are academically selective so one way of interpreting this is that it reflects a negative association between academic ability and cannabis use, consistent with McVicar and Polanski (2012).

[19] Unlike in the US, in Northern Ireland these are generally non-selective public schools that serve Catholic communities, which helps explain why we find a positive association with cannabis use where Gaviria and Raphael (2001), using US data, find a negative association with drug use.

use and peer cannabis use, with a one percentage point increase in the proportion of the school grade that use cannabis associated with a .69 percentage point increase in own probability of cannabis use. Note also that the overall explanatory power of the regression increases when peer behaviour is included compared to the controls-only version of (1). As we would expect, the magnitude of the estimated peer effect falls slightly when postcode dummies are included in (1) to control for neighbourhood-level correlated effects – these are jointly significant, with a p-value of .004 – but we still obtain a large, highly significant and positive association between own use and reference group use.

<Table 3 around here>

Table 3 also presents the IV estimate of the peer effect for the model including postcode dummies and contextual effects. As for the OLS estimate the suggestion is that peer effects in early adolescent cannabis use are large, positive and highly statistically significant, with a one percentage point increase in cannabis use prevalence in the school grade leading to a .81 percentage point increase in own probability of cannabis use. The contextual effects are jointly insignificant, with a p-value of .80, lending support to the Gaviria and Raphael (2001) argument, and lending support to our OLS estimates based on this assumption.

Note that the IV estimate is *larger* (although not significantly) than the equivalent OLS estimate, which although not an uncommon finding in the peer effects literature (e.g. Gaviria and Raphael, 2001; Lundborg, 2006; McVicar, 2012), is somewhat counterintuitive. If we are prepared to assume that this IV strategy is valid, then the implication is either that negative biases (e.g. negative simultaneity bias, attenuation bias due to measurement error) outweigh positive biases in the OLS estimates, or that the variation in peer cannabis use engendered by the instrument occurs disproportionately in the reference groups of individuals who are more than usually sensitive to peer influence.[20]

That *both* the OLS and the IV estimates are at the higher end of the range of broadly comparable estimates in the existing literature, which generally fall between .1 and .6, could be related to the fact that we estimate peer effects for cannabis use in the last 12 months, which is more likely to pick up peer effects to and from 'one-time-tryers' or very occasional users than studies that use measures of cannabis use over a shorter period. When we repeat the Table 3 estimations using monthly use in place of use in the last 12 months (for both the

---

[20] If peer effects are heterogeneous, then IV provides a *local average treatment effect* (LATE) estimate of peer effects rather than the average treatment effect (ATE) estimate (Lefgren, 2004).

individual and the reference group), the estimates are smaller in magnitude in all cases, although still generally large, positive and statistically significant at standard levels.[21] Further, our age group – 13/14 year olds, i.e. early adolescents – may be particularly sensitive to peer influences.[22] There may also be cultural or other contextual differences between Belfast adolescents and adolescents in other studies, e.g. the Add Health study for the US, that impact on the magnitude of peer effects.[23]

5.2. Treating Nominated Friends as the Reference Group

Next consider estimates of peer effects where friends are taken as the reference group (equation (2)). Results are presented in Table 4. Estimating (2) by OLS suggests a strong, highly significant and positive association between own cannabis use and friends' cannabis use, with a one percentage point increase in the proportion of friends that report using cannabis in the last 12 months associated with a .6 percentage point increase in own probability of reporting cannabis use in the last 12 months. The magnitude of this estimate falls to .55 when we include school fixed effects to control for common school-level factors.[24] Note that in both cases the magnitude of the estimated peer effect appears similar to that for peer effects from grade-mates.

Unlike in (1) where the school grade is treated as the reference group, we can also include contextual effects in (2) – averages of the *X*s or proportions across the set of nominated friends – and still arguably get close to identifying the peer effect from friends' cannabis use using OLS, because each individual has a difference reference group of nominated friends (see Calvo-Armengol et al., 2009). When we do so the peer effect estimate falls to .49, but remains statistically significant at the 99% level. The contextual effects are also jointly significant in this case, unlike in the schoolmates case, with a p-value of .01.

<Table 4 around here>

---

[21] Results are available on request.

[22] For example, Berndt (1982) and Cook et al. (2007) suggest that young people have especially strong social identity needs at this time.

[23] For a general argument see McVicar (2011). For a description of the very particular post-conflict context for drug use in Northern Ireland see McEvoy et al. (1998).

[24] This drop is robust to replacing school dummies with postcode dummies or including both school and postcode dummies.

Table 4 also presents two IV estimates. The first, using the proportion of friends who report having older non-school friends as the instrument, again suggests a large, positive, and highly statistically significant peer effect, with a one percentage point increase in cannabis use prevalence in among nominated friends leading to a .75 percentage point increase in own probability of cannabis use. As in the grade-mates case, the IV estimate is larger in magnitude than the equivalent OLS estimate (and the difference is larger). Again, there are legitimate reasons why this might be the case as discussed above, but the more likely explanation is that the instrument is less excludable in the case of friends than in the case of grade-mates. The joint significance of the contextual effects supports this interpretation.

The second IV estimate presented in Table 4 addresses this concern over the standard-approach IV by using the characteristics of friends of nominated friends who are not themselves ego's friends – friends once removed – to instrument for friends' cannabis use. Once again the suggestion is of large, positive peer effects, although in this case the increased imprecision of the estimated coefficient, reflecting the lower first-stage F-statistic, renders it statistically insignificant. This second IV estimate is much closer to the equivalent OLS estimate, with the point estimate of .40 now *below* the equivalent OLS estimate of .49. In other words, our preferred IV strategy, coupled with school fixed effects and controls for contextual effects, does in fact suggest a positive bias in the equivalent OLS estimate.

We also repeat the Table 4 estimations using *monthly use* instead of *use in the last 12 months*. As for peer effects from grade-mates, the estimates are generally smaller in magnitude, although still large, positive and statistically significant at standard levels.[25]

How do these estimates compare with existing estimates of peer effects in cannabis use between friends? Unfortunately, unlike in the grade-as-reference-group case, there are few friends-based estimates in the existing literature with which these estimates can be directly compared. A partial exception is Eisenberg (2004) who reports considerably smaller peer effects for cannabis use using Add Health data. Otherwise, where peer effects between friends have been previously estimated in the literature, they either use lagged peer behaviour in place of current peer behaviour (Clark and Lohéac, 2007), perceived behaviour of friends rather than actual (or at least self-reported) behaviour (e.g. Kawaguchi, 2004; McVicar and

---

[25]Results are available on request.

Polanski, 2012), or estimate peer effects for different substance-use behaviours (e.g. Norton et al., 2003).[26]

5.3. Are There Peer Effects from 'Non-friends'?

Our results so far point to large, positive and generally statistically significant peer effects in cannabis use, regardless of whether nominated friends or grade-mates are treated as the reference group. Is it a coincidence that the magnitude of the estimated peer effect appears to be similar for both reference groups, at least according to the preferred OLS estimates? Our conjecture, building on the earlier findings of McVicar and Polanski (2012), is that it is not. On the contrary, if grade-mates that are not nominated as friends are an irrelevant reference group for ego's own cannabis use, as suggested by McVicar and Polanski (2012), then the grade-level cannabis use variable and the nominated friends' cannabis use variable are essentially capturing the same reference group behaviour, i.e. friends within the grade.

A numerical example can help to clarify matters here. Consider a school grade with 101 students (ego plus 100 grade-mates), and assume the peer effect estimate between friends is ϕ, and the peer effect estimate between non-friends is π. Imagine ego nominates 10 of the 100 grade-mates as friends, leaving 90 as 'non-friends'. Now imagine one grade-mate becomes a cannabis user at random. If this grade-mate is a friend this corresponds to a 10 percentage point increase in the prevalence of cannabis use among nominated friends, and the associated impact on own probability of use is 10ϕ. If the grade-mate is a non-friend, this corresponds to a 100/90 percentage point increase in the prevalence of cannabis use among non-friends, with an associated impact on own probability of use of $100\pi/90$. Finally, because the probability that this grade-mate is a friend (non-friend) is .1 (.9), the expected impact of the change in behaviour of a random grade-mate is .1*10ϕ + .9*$100\pi/90$, i.e. $\phi + \pi$.[27] If we assume that non-friends have no influence on own behaviour, then $\pi=0$ and the expected impact of a one grade-mate change in behaviour is the same whether we treat the entire grade as the reference group or just nominated friends as the reference group.

---

[26] The closest comparator otherwise is probably Norton et al. (2003), which presents OLS estimates of peer effects in tobacco smoking from up to three nominated 'close friends' within school which are similar in nature to those presented here. The study omits school fixed effects and contextual effects, however, and presents no IV estimates in addition to the OLS estimates. Using this approach, they estimate coefficients of around .53 on friends' behaviours, which is close to our own OLS estimate for cannabis use.

[27] Note that this reflects the fact that reference group behaviour is expressed as a proportion. If reference group behaviour is expressed in terms of the number of individuals that use cannabis, then the grade-mate peer effect is a weighted average of the friends and non-friends peer effects.

To investigate this question further we use the information on nominated friends to split each individual's grade-mates into friends and non-friends, treating the two groups as separate reference groups. We then estimate the impact of behaviour in both reference groups on own behaviour, i.e. equation (3). Results are presented in Table 5. There are three sets of OLS estimates: first, without postcode dummies on the full sample; second, on the sample restricted to those with postcode information but without postcode dummies; and third, including postcode dummies.[28] None of the three variants of the model suggests a strong peer effect from non-friends. In the first case the estimated peer effect for non-friends is just .13 and is only marginally statistically significant. When postcode dummies are included the coefficient falls to zero. This zero or close to zero association between own cannabis use and non-friends cannabis use contrasts with the large, positive and highly significant association between own use and friends' use. All three estimates are close to .6, as is the case in Table 4 where non-friends' behaviour is omitted.

<Table 5 around here>

As before, we are still concerned that OLS estimates of peer effects from friends may be biased due to selection and reflection. The same is potentially true for non-friends, although both sources of potential bias may be weaker in the case of the non-friends peer effects estimate. Table 5 therefore also presents IV estimates for (3), using the same proportion of friends reporting older non-school friends instrument as before along with the equivalent proportion of non-friends reporting older non-school friends instrument for non-friends. Once again the estimated peer effect from friends remains large, positive and statistically significant, while the peer effect from non-friends remains close to zero and statistically insignificant. Also note that the evidence of zero peer effects from non-friends presented in Table 5 supports the friends-once-removed IV strategy used earlier to estimate (2).

6. Conclusions

This paper presents micro-econometric estimates of peer influences in early adolescent cannabis use using school-based survey data not previously exploited for this purpose. The

---

[28] As for the estimates where the school grade is treated as the reference group, we cannot include school dummies in (3) because school dummies are collinear with a combination of the friends and non-friends cannabis use variables. But if we assume no peer effects from non-friends, then including non-friends' use in (3) similarly controls for school-level factors that influence cannabis use.

data set has a number of attractive properties allowing us to make contributions to the substance use peer effects literature on three fronts.

First, we present credible IV-neighbourhood fixed effects estimates of peer effects – arguably less susceptible to instrument validity concerns than at least some previous IV estimates in the literature – where the school grade is treated as the reference group. Our results suggest the existence of large, positive, highly statistically significant peer effects in cannabis use from this reference group, whether peer behaviour is instrumented or not.

Second, we exploit information on nominated friends in the data to estimate peer effects in cannabis use where friends are treated as the relevant reference group. Ours is one of only a handful of studies that provides such estimates, and we argue that our estimation approach – in particular using information for friends of friends who are not nominated friends of the individual to instrument for nominated friends' behaviour, together with school fixed effects – offers the best chance of correctly identifying peer effects in cannabis use from this select group of studies. Again, our results suggest the existence of large peer effects from this reference group.

The main contribution of the paper, however, is to show that friends are the relevant reference group for adolescent cannabis use, and schoolmates who are not nominated friends have no influence on own cannabis use. This is the first paper to do so conclusively. We argue that this is a critical distinction for parents concerned with minimising exposure to negative peer influences. Further, the suggestion is that policy interventions that aim to exploit social multipliers to reduce adolescent substance use may be more successful at the friendship cluster level rather than school grade level.

References

Berndt, T.J. (1982). 'The features and effects of friendship in early adolescence.' *Child Development,* 53, 1447-1460.

Calvo-Armengol, A., Pattachnini, E. and Zenou, Y. (2009). 'Peer effects and social networks in education.' *Review of Economic Studies*, 76, 1239-1267.

Clark, A.E. and Lohéac, Y. (2007). 'It wasn't me, it was them! Social influence in risky behaviour of adolescents.' *Journal of Health Economics*, 26, 4, 763-84.

Cook, T., Deng, Y., and Morgano, E. (2007). 'Friendship influences during early adolescence: the special role of friends' grade point average.' *Journal of Research on Adolescence*, 17, 2, 325-356.

Eisenberg, D. (2004).'Peer effects for adolescent substance use: do they really exist?' Working Paper UC-Berkeley School of Public Health.

Fletcher, J.M. (2010). 'Social interactions and smoking: evidence using multiple student cohorts, instrumental variables and school fixed effects.' *Health Economics*, 19: 466-484.

Gaviria, A. and Raphael, S. (2001). 'School-based peer effects and juvenile behaviour.' *Review of Economics and Statistics*, 83, 257-68.

Hoxby, C.M. (2000). 'Peer effects in the classroom: learning from gender and race variation.' NBER Working Paper No. 7867.

Kawaguchi, D. (2004). 'Peer effects on substance use among American teenagers.*' Journal of Population Economics*, 17, 351-67.

Krauth, B.V. (2005). 'Peer effects and selection effects on smoking among Canadian youth.' *Canadian Journal of Economics* 38, 735-756.

Krauth, B.V. (2007). 'Peer and selection effects on youth smoking in California.' *Journal of Business and Economic Statistics*, 25, 288-98.

Kuepper, R., van Os, J., Lieb, R., Wittchen, H.-U., Hofler, M., Henquet, C. (2011). 'Continued cannabis use and risk of incidence of psychotic symptoms: 10 year follow-up cohort study.' *British Medical Journal*, 342, 738-745.

Lefgren, L. (2004). 'Educational peer effects and the Chicago public schools.' *Journal of Urban Economics*, 56, 169-191.

Lundborg, P. (2006). 'Having the wrong friends? Peer effects in adolescent substance use.' *Journal of Health Economics*, 25: 214-233.

Manski, C. (1993). 'Identification of endogenous social effects: the reflection problem.' *Review of Economic Studies*, 60, 531-42.

Manski, C. (2000). 'Economic analysis of social interactions.' *Journal of Economic Perspectives*, 14: 115-36.

McCrystal, P., Percy, A. and Higgins, K. (2007). 'Frequent cannabis use among 14/15 year olds in Northern Ireland.' *Drug and Alcohol Dependence*, 88, 19-27.

McEvoy, K., McElrath, K. And Higgins, K. (1998). 'Does Ulster still say no? Drugs, politics, and propaganda in Northern Ireland.' *Journal of Drug Issues*, 28, 1.

McVicar, D. (2011). 'Estimates of peer effects in adolescent smoking across twenty six European countries.' *Social Science & Medicine*, 73: 1186-1193.

McVicar, D. (2012). 'Cross-country estimates of peer effects in adolescent smoking using IV and school fixed effects.' Working Paper 7/12, Melbourne Institute of Applied Economic and Social Research.

McVicar, D. and Polanski, A. (2012). 'Peer effects in UK adolescent substance use: never mind the classmates?' Working Paper 8/12, Melbourne Institute of Applied Economic and Social Research.

Norton, E.C., Lindrooth, R.C. and Ennett, S.T. (2003). How measures of perception from survey data lead to inconsistent regression results: evidence from adolescent and peer substance use.' *Health Economics*, 12, 139-48.

Patton, G.C., Coffey, C., Carlin, J.B., Degenhardt, L., Lynskey, M. and Hall, W. (2002). 'Cannabis use and mental health in young people: cohort study.' *British Medical Journal*, 325 (7374), 1195+

Powell, L.M., John, A. and Hana, R. (2005). 'The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behaviour.' *Journal of Health Economics*, 24, 950-68.

Sen, A. (2009). 'Estimating the impacts of household behaviour on youth smoking: evidence from Ontario, Canada.' *Review of Economics of the Household* 7: 189-218.

Soetevent, A.R. and Kooreman, P. (2007). 'A discrete choice model with social interactions: an application to high school teen behaviour.' *Journal of Applied Econometrics* 22: 599-624.

Van Ours, J.C. and Williams, J. (2009). 'Why parents worry: initiation into cannabis use by youth and educational attainment.' *Journal of Health Economics*, 28, 1, 132-142.

Vasquez, B.E. (2010). 'Methodological difficulties in modelling peer influence.' *Social Science Research*, 39, 6, 950-962.

Table 1: Observed Variables and Sample Means

Variable	Mean
Own use of cannabis in the past 12 months	.28
Uses cannabis at least monthly	.11
Sex-male	.47
Age at which began school, years	11.7
Number of friends nominated	7.35
Number of friends also valid	6.52
Number of brothers	1.03
Number of sisters	.98
Number of cars in household	1.46
Lives with single parent	.15
Lives with parent and step-parent or parent's partner	.08
Lives in alternative family structure	.02
Family structure missing	.01
Wants to leave school at 16	.20
Wants to leave school at 16 missing	.03
Has older friends in the school	.23
Has older friends outside of school	.22
Mother works full-time	.33
Mother works part-time	.29
Mother work missing	.10
Father works full-time	.63
Father works part-time	.08
Father work missing	.20
Attends single-sex school	.61
Attends grammar school	.45
Attends Catholic school	.44
Nobs	4,417

Notes: All wave 3 respondents other than those with missing cannabis use information, age information, pupils in alternative education.

Table 2: Baseline Model for Cannabis Use in Previous 12 Months

	OLS Coefficient	Robust Standard Error
Sex-male	.071***	.020
Age	.062***	.017
Number of classmates (*100)	.019	.023
Number of brothers	.013	.008
Number of sisters	.013	.008
Number of cars	.006	.007
Single parent family	.103***	.028
Step family	.128***	.030
Alternative family	.056	.043
Want to leave school at 16	.136***	.023
Want to leave school at 16 missing	.229***	.044
Older School Friends	.115***	.017
Older Non-School Friends	.116***	.016
Mother works full-time	.033*	.019
Mother works part-time	.023	.023
Father works full-time	-.019	.026
Father works part-time	.005	.028
Attends single-sex school	-.012	.031
Attends grammar school	-.103***	.028
Attends Catholic-managed school	.059**	.027
Nobs	4,417	
R-Squared	.121	

Notes: *** Significant at 1% probability level, **5% level, *10% level. Additional control variables are dummies for missing values for number of brothers, number of sisters, household structure, father works missing, mother works missing (none of which have a statistically significant impact on cannabis use).

Table 3: Peer Effects in Cannabis Use, Classmates as Reference Group

	Coefficient	Robust standard error	Nobs	R-squared
A) OLS	.693***	.071	4417	.141
B) OLS, sample with postcode data	.719***	.080	3533	.126
C) OLS with postcode dummies	.620***	.103	3533	.135
D) IV with postcode dummies and contextual effects	.814**	.373	3533	.137

Notes: *** Significant at 1% probability level, **5% level, *10% level. Models include the controls as listed in Table 2. Missing postcode data means we can only specify area dummies for a subsample. The IV models use the proportion of grade-mates who have older non-school friends as a single instrument. The F-statistic for the significance of the excluded instrument in the first stage regression is 441.

Table 4: Peer Effects in Cannabis Use, Friends as Reference Group

	Coefficient	Robust standard error	Nobs	R-squared
A) OLS	.602***	.026	4289	.232
B) OLS, with school fixed effects	.553***	.029	4289	.237
C) OLS, with school fixed effects and contextual effects	.485***	.032	4289	.245
D) IV with school fixed effects and contextual effects	.749***	.167	4289	.229
E) IV with school fixed effects, friends once removed, and contextual effects	.397	.306	4276	.243

Notes: *** Significant at 1% probability level, **5% level, *10% level. Standard errors are clustered at the school level. Models include the controls as listed in Table 2. The IV models use the proportion of nominated friends who have older non-school friends as a single instrument. IV 'once removed' uses the same instrument but for friends once removed. F-statistics for the significance of the excluded instrument in the first stage regression are 72 and 11 respectively in the two versions of the IV model.

Table 5: Peer Effects in Cannabis Use, Friends & Non-friends

	Friends' cannabis use, coefficient (robust standard error)	Non-friends' cannabis use, coefficient (robust standard error)	Nobs	R-squared
A) OLS, full sample	.590***	.126*	4289	.232
	(.027)	(.070)		
B) OLS, sample with postcode data	.610***	.118	3454	.221
	(.031)	(.080)		
C) OLS with postcode dummies, friends & non-friends	.598***	-.004	3454	.228
	(.032)	(.103)		
D) IV with postcode dummies and contextual effects	.530**	.082	3454	.240
	(.224)	(.383)		

Notes: *** Significant at 1% probability level, **5% level, *10% level. Models include the controls as listed in Table 2. Missing postcode data mean we can only specify area dummies for a subsample. The IV models use the proportion of friends and non-friends who have older non-school friends as a single instrument respectively in each first stage equation.