University of Neuchatel

Institute of Economic Research

IRENE, Working paper 13-03

The impact of cannabis use on short-term educational outcomes

Joachim Marti and Lionel Perini***

** Yale University*

*** University of Neuchatel*

UNIVERSITÉ DE NEUCHÂTEL

Institut de recherches économiques

The impact of cannabis use on short-term educational outcomes

Joachim Marti[a)] and Lionel Perini[b)]

a) Joachim Marti, PhD
Postdoctoral Fellow
Yale School of Public Health
Department of Health Policy and Management
60 College Street
New Haven, CT 06520
USA

Email: jomswiss@gmail.com or joachim.marti@unine.ch
Telephone: +41 76 322 95 33
Fax: +1 203 785 6287

b) Lionel Perini
Teaching and Research Assistant
University of Neuchâtel
Institute for Research in Economics
Pierre-à-Mazel 7
2000 Neuchâtel
Switzerland

Email: lionel.perini@unine.ch

The impact of cannabis use on short-term educational outcomes

Abstract

In this paper we use longitudinal data on Swiss adolescents to investigate the impact of cannabis use on short-term educational performance. We focus our analysis on high school students and analyze various outcomes, including absenteeism, grades, and motivation. We exploit the panel nature of the data and control for a rich set of individual and family characteristics measured at the end of compulsory school. Results from both fixed effects regressions and propensity score matching indicate that high school students who smoke cannabis skip on additional half day of school per month and are 15-20% more likely to obtain poor grades. In addition, our empirical approaches highlight the importance of taking unobserved heterogeneity into account when assessing the impact of substance use on education.

Keywords: cannabis, education, adolescents, human capital, propensity score matching

JEL codes: I12, I18, I21, C23

1. Introduction

Data from the recent Addiction Monitoring Survey conducted in Switzerland reveal particularly high levels of cannabis consumption in the country, especially among adolescents and young adults. The prevalence of past 30 days use almost reaches 10% in the 15-24 age group, and nearly 50% of young adults aged 20-24 reports having smoked cannabis at least once in their lifetime [1]. Additionally, approximately one-fifth of adolescent and young adult consumers report smoking cannabis daily or almost daily [1]. The high levels of cannabis consumption observed in the country may be partly explained by a lack of understanding of its harmful consequences. A qualitative study conducted in various age groups in Switzerland shows that, overall, people do not have a very clear perception of the risks of cannabis use and of what constitutes misuse [2]. The authors call for a more developed prevention approach, including better provision of information about the risks of consumption. This is especially important considering the fairly permissive legislation of cannabis use in Switzerland that may send confusing signals about risks and social norms. Formally, the product is illicit but consequences are limited for consumers and the law is enforced with various degrees of intensity in the country.

A consequence of cannabis use that has attracted increasing attention is its potential impact on cognitive abilities and, ultimately, on the accumulation of human capital among youths. International evidence shows that adolescents are both using cannabis at younger ages and that early initiation often corresponds to worse cognitive outcomes [3-5]. In a study that assess cognitive performance of marijuana users, Gruber et al. (2012) [4] conclude that exposure to the product during adolescence affects brain development and finds that age of onset, frequency of consumption and level of consumption influence the strength of this relationship. Fontes et al. (2011) [3] find similar results and emphasize the particularly detrimental effects of early onset. In a recent study using longitudinal data from New Zealand, Meier et al. (2012) [5] compare the evolution of cognitive functioning before and after cannabis initiation and show worse deteriorations in outcomes among early and persistent users.

This impact on cognitive skills may impair student concentration and learning ability. Equally concerning is the time that student may spend using cannabis instead of attending classes, or the detrimental impact on motivation and involvement that the product may have. In other words, cannabis use could reduce the ability to learn and successfully complete studies through reduced school attendance [6] or "classroom presenteeism."

Many studies have found evidence of an association between cannabis use and poor schooling outcomes [7-11] but only a few employ empirical strategies that address potential bias arising from reverse causality (e.g., psychological distress due to school difficulties may increase the perceived benefits of consumption) or unobserved heterogeneity (unobserved individual characteristics, such as time preferences, that are both influencing cannabis use and schooling outcomes) [6, 11-15]. For instance, Van Ours and Williams (2009) [15] focus on the impact of age at initiation on dropout rates among Australian adolescents. Using bivariate duration models, they show that early onset of cannabis use has a detrimental impact on years of education completed and significantly increase the likelihood of school dropout. McCaffrey (2010) [12] find evidence of an impact of heavy and persistent cannabis use on high school dropout using propensity score matching. While they argue that their results are probably driven by time-varying unobserved heterogeneity rather than by effects on cognitive abilities, the mechanisms remain unclear.

These studies mostly focus on educational attainment outcomes, such as drop out rates or the number of years of education completed. Intermediate (short-term) outcomes are rarely considered, leaving underlying mechanisms poorly understood. Only little econometric evidence exists on the impact of cannabis use on outcomes such as learning ability, grades or motivation. Notable exceptions are papers by Pacula et al. (2003) [13] and Roebuck et al. (2004) [6]. The former provides evidence of a negative impact of cannabis use on standardized test scores but show that the estimated effects considerably shrink after accounting for unobserved heterogeneity. The latter uses religiosity at the individual level as an instrument for cannabis use and show that cannabis users skip more school days than non-users.

In this paper, we build on this body of work and investigate the pathways through which cannabis use may affect educational attainment among high school students. We use data from a longitudinal dataset of adolescents in Switzerland and focus on several short-term academic outcomes, including concentration problems, learning difficulties, absenteeism and poor grades. We compare results from individual fixed effects models and propensity score matching and assess the sensitivity of the latter to potential unobserved heterogeneity using Rosenbaum bounds [16]. We control for several usually unobserved personality traits such as persistency and self-esteem, and for a rich set of family and individual characteristics measured at completion of compulsory education. We find consistent evidence that cannabis use reduces school attendance and increases the likelihood of poor educational performance among high school students. However, our empirical results highlight the importance to account for unobserved heterogeneity in the substance use-education relationship.

2. Data

Our data come from the Swiss Transition from Education to Employment (TREE) survey[1], which is nationally representative and longitudinal. TREE collects information annually, starting in 2001, on a series of education, work, and health-related variables, along with rich information on psychological traits. The baseline TREE sample consists of a subsample of 5,528 adolescents who responded to the OECD Program for International Student Assessment (PISA) questionnaire in 2000, which takes place at the end of compulsory schooling (i.e., at 15 years old for the vast majority of students). We are able to match the TREE survey information with the PISA responses and therefore have access to a wide variety of background characteristics for each respondent. This baseline data includes information such as family characteristics (including educational support), intermediate school quality indicators and measures of cognitive ability (e.g., reading and math test scores).

[1] The Swiss youth panel study TREE (Transitions from Education to Employment; tree.unibas.ch) has been running since 2000 and has since been funded by the Swiss National Science Foundation, the University of Basel, the Swiss Federal Office of Statistics, the Federal Office of Professional Education and Technology, and the cantons of Berne, Geneva and Ticino. Distribution: Dataservice, FORS, Lausanne. The dataset is available to all interested researchers and can be ordered at the Data Archive of the Swiss Foundation for Research in Social Science (FORS) in Lausanne, Switzerland.

In Switzerland, a majority of adolescents are involved in professional tracks after compulsory school (i.e. vocational school, apprenticeship). Because these educational programs differ widely in terms of academic content and study hours, we focus on high-school students (i.e., adolescents enrolled in an academic maturity school), a more homogenous population that follows a full-time education program. Additionally, because our identification strategy relies on individual-level changes in cannabis use and because we use lagged cannabis use in our models, we restrict the sample to students that were observed for each year between 2001 and 2003 (N=1,416). It is worth noting that we could have included a fourth wave. However, the duration of studies varies from 3 to 4 years across regional (cantonal) systems. Therefore, students in 2004 may be a significantly different selected group than students in their first three years of high school. We dropped respondents with incomplete information on cannabis use and on other control variables and obtain an analysis sample of approximately 1,100 high school students. The exact size of each analysis sample depends on the outcome under investigation and on the specification used.

In our analysis, we focus on six short-term outcomes that measure different aspects of schooling. First, we consider absenteeism, which is defined as the number of days the student was absent from school during the previous month. Second, we create an index for school difficulties whose values range from 0 to 5. More precisely, the index counts the number of questions, among the following five, to which the respondent answered “often” or “very often”: “If I don’t study during the weekend, I can hardly satisfy school requirements”, “I have too much work at school”, “I can hardly manage the amount of homework”, “The subjects of the lessons change so fast, that I have trouble to keep up” and “At school I often feel out of my depth”. Next, we create two binary indicators reflecting lack of engagement/concentration and engagement/motivation that equal one if the respondent answers “no” to the following statements, respectively: “I work very concentrated at school” and “Usually I am fully present at school.” Then, we create a binary outcome *Poor grades* which takes the value one if the student mentions having had at least one failing grade in her last grade report. Our last outcome is a binary indicator of recent concentration problems (“Over the last month, did you suffer from lack of concentration?”). Although we are estimating reduced form equations for each of these outcomes, it is conceptually important to distinguish between performance outcomes and mechanisms. Our performance outcome is *Poor grades*, while all other outcomes pertain to the education production function itself: exposure to education (school days), concentration, motivation, and learning ability. It is beyond the scope of this paper to estimate a full structural education production function.

Our main variable of interest is the frequency of cannabis use. The questionnaire asks about the frequency of consumption over the month preceding the interview with possible answers ranging from "never" to "daily use" (i.e., never, 1-3 times a month, 1-2 times a week, 3-5 times a week, and daily). We construct two dummy variables. First, we create an indicator for any use that takes the value 1 if the individual has smoked cannabis at least once during the month preceding the interview. Then, we create an indicator for frequent cannabis use that makes the distinction between frequent users (at least once a week) and never- and occasional users (i.e., never, 1-3 times a month).

The PISA survey includes an extensive set of individual and family characteristics (i.e., gender, living in a nuclear family, parental education, parental wealth, parental socio-economic status and number of siblings). It also collects information on household educational support (i.e., parental educational support, number of books at home and educational resources at home) and on educational outcomes during the last year of secondary school (i.e., language and math test scores). We were able to match each respondent to its information collected in the PISA 2000 survey and therefore obtain a rich set of baseline (i.e., pre high-school) relevant characteristics.

In addition, the TREE survey itself includes a large set of variables measuring psychological traits, non-cognitive skills and substance use. We exploit this information and use a series of scales measured at Wave 1 (i.e., in 2001) reflecting persistency, self-efficacy, self-esteem and positive attitude. Each of these psychological variables is constructed by aggregating answers to a series of questions (details on the construction of these variables are provided in the Appendix). We also create dummy variables for alcohol and tobacco consumption at Wave 1. Summary statistics for all relevant variables are presented in Table 1.

3. Empirical approach

Our objective is to uncover the impact of cannabis consumption on a series of short-term educational outcomes. The main empirical challenge is that any observed correlation between cannabis use and poor educational outcomes may be due to the influence of common unobserved factors; or it may be that low performance at school increases the propensity to engage in risky behaviors. In this paper, we exploit the longitudinal nature of our data and a rich set of control variables to overcome these potential issues. Precisely, we compare results from individual fixed effects regressions with those obtained using a non-parametric approach (i.e., propensity score matching). The use of two empirical strategies allows us to assess the robustness of our results and provides different ways to evaluate the importance of unobserved heterogeneity.

An alternative option to uncover the causal impact of cannabis use on short-term academic performance would have been to use an instrumental variable approach. However, credible instruments are challenging to find in substance use research [17], especially in the case of illicit drugs. A variable that credibly impacts academic performance only through cannabis use is not available in our case; we therefore rely on changes in consumption over time for identification.

Pooled OLS and fixed-effects

We start by estimating a series of OLS and linear probability models to investigate the association between cannabis use and educational outcomes. More precisely, we model the association between lagged cannabis use, $c_{i,t-1}$, and contemporaneous educational outcomes, y_{it}. Our baseline specification is:

$$y_{it} = \alpha_0 + \alpha_1 c_{i,t-1} + \alpha_2 x_i + \delta_t + \varepsilon_{it} \qquad (1)$$

where x_i represents a vector of baseline characteristics, δ_t is a wave indicator that accounts for trends, and ε_{it} is an idiosyncratic error term. We use a lagged measure of cannabis use in order to mitigate potential bias arising from reverse causality, even if some evidence from Spain suggests that school failure does not impact cannabis use [18]. Also, it is worth noting that for the outcomes "engagement", "motivation", "poor grade" and "concentration", we use linear probability models whereas these variables are dichotomous. We decided to estimate linear probability models for ease of interpretation better comparability between outcomes.

For the OLS results to be considered as unbiased, we must make the assumption that lagged consumption is exogenous. However, it is likely that some unobserved individual characteristics affect both consumption and the outcomes of interest (e.g., time preferences, peer influence, rebelliousness or preference for deviant behavior). We therefore exploit the longitudinal nature of our data and extend (1) by controlling for individual fixed effects, η_i:

$$y_{it} = \alpha_0 + \alpha_1 c_{i,t-1} + \alpha_2 x_i + \delta_t + \eta_i + \varepsilon_{it} \tag{2}$$

Practically, we use the within-estimator that purges the estimates from the influence of time-invariant individual characteristics:

$$(y_{it} - \overline{y}_i) = \gamma_1 (c_{i,t-1} - \overline{c}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i) \tag{3}$$

Our main identifying assumption is that after controlling for fixed individual characteristics, there are no other unobserved factors that both influence lagged cannabis use and educational outcomes. Naturally, some unobserved time varying individual factors may still bias our results, but our focus on a relatively short time period mitigates this concern. In all models, standard errors are clustered at the individual level.

Propensity score matching

To assess the robustness of our results, we use propensity score matching (PSM), a semi-parametric approach that relaxes the linearity assumption inherent to the use of OLS and FE estimators. Precisely, we compare short-term educational outcomes of cannabis users to those of a matched group of non-users (or occasional users) with similar observed characteristics. Due to dimensionality issues, performing exact matching with a large number of covariates is challenging [19]. PSM overcomes this problem by matching individuals based on their estimated probability to belong to the treatment group (i.e., their propensity score) [20]. In our approach, we exploit the longitudinal nature of our data and define treated adolescents as those who report smoking cannabis at any frequency at Wave 2 but who did not smoke at Wave 1. In other words, our treatment of interest is the onset of cannabis use between waves (in alternative specifications, we modify the treatment of interest and focus on the onset of frequent cannabis use).

With this procedure, we compare individuals that are similar in terms of observed characteristics but that differ in their use of cannabis. Identification relies on the assumption that there are no remaining unobserved characteristics correlated with both cannabis onset and educational outcomes. While we are not able to formally test it, we perform below a sensitivity analysis to assess the robustness of our results to this assumption.

We start by estimating the probability of cannabis initiation, i.e., the probability to belong to the treatment group. We use pre-treatment characteristics as defined above, measured both in PISA and at Wave 1, that are likely to influence both cannabis use and education and estimate logit models of the form:

$$p(x_i) = p(D_i = 1 \mid x_i) = F(\beta x_i + \gamma u_i) \quad (4)$$

where D_i is the treatment variable (i.e., the participation decision) that indicates whether the individual started to smoke cannabis (regularly) at Wave 2, x_i is a vector of pre-determined characteristics, u_i represents unobserved heterogeneity, and F is the logistic distribution. The parameter γ reflects potential correlation remaining between unobserved characteristics and the participation decision. We first assume conditional independence, which implies that, after controlling for x_i, γ is equal to zero. Below, we gauge the sensitivity of our results to the conditional independence assumption (CIA).

The next step consists in forming pairs of treated and untreated individuals that have similar predicted probability to be treated (i.e., a similar propensity score). We use several matching estimators, including kernel matching and bias-corrected nearest-neighbor matching with single and multiple (i.e., 5) neighbors [21]. We then estimate the *average treatment effect on the treated* (ATT) by comparing educational outcomes between the two groups. We have, for each outcome k:

$$\tau_{ATT}^{k} = E\left[Y_{1i}^{k} \mid D_i = 1, p(x_i)\right] - E\left[Y_{0i}^{k} \mid D_i = 0, p(x_i)\right] \text{ with } k \in \{1,...,6\} \qquad (5)$$

In order to evaluate the sensitivity of our results to potential unobserved heterogeneity, we use Rosenbaum bounds [16]. This method examines how the confidence intervals around the ATT are affected by different assumptions about the value of γ in (3). To provide intuition about this procedure, consider two individuals from a matched pair, indexed by *i* and *j*, who have the same values of observed covariates. Rosenbaum (2002) [16] has shown that, in the presence of unobserved characteristics that affect the participation decision, these two individuals may differ in their odds of receiving treatment by a factor Γ (see also Rosenbaum, 2003, 2005) [22-23]:

$$\frac{1}{\Gamma} \leq \frac{p_i(1-p_j)}{p_j(1-p_i)} \leq \Gamma, \text{ with } \Gamma \geq 1 \qquad (6)$$

If the (untestable) conditional independence assumption holds, Γ is equal to one. The sensitivity analysis proposed by Rosenbaum computes the range of significance levels for several values of Γ and therefore informs us about how sensitive our findings are to potential biased treatment assignment[2]. This approach does not provide a formal test of the CIA but allows researchers to gauge the sensitivity of their findings to potential selection on unobservables.

4. Results

OLS and FE

Table 2 provides results for any cannabis use (panel A) and frequent cannabis use (panel B). For each outcome (i.e., absenteeism, school difficulties, engagement, motivation, poor grades and concentration), the table shows the coefficient of interest obtained with both the OLS and FE specifications (tables with full results are presented in Appendix B).

[2] It is worth noting that we make the assumption of a potential positive selection bias (i.e., unobserved factors that are positively correlated with both cannabis use and poor educational outcomes).

OLS results for absenteeism suggest that cannabis use increases the number of school days skipped among high school students, irrespective of frequency of use. After controlling for fixed unobserved factors the coefficient remain significant for any use only: fixed-effects results show that cannabis users skip on average 0.6 additional school days per month as compared to non-users. We do not find consistent evidence of an impact of cannabis use on the index of self-reported school difficulties, except in the OLS model for frequent use. Frequent cannabis use has a positive impact on self-assessed lack of attention in the classroom. Fixed-effects results indicate that frequent users are approximately 13% more likely to report attention deficit in the classroom. Models for motivation do not suggest any association between use and this outcome.

The most sizeable effects of cannabis use are found for the outcome *Poor grade*. Fixed-effects estimates are positive and significant in all models. Overall, results suggest that cannabis consumption increases the probability of receiving poor grades in the last grade report by approximately 10 to 20 percentage points, with stronger effects found among frequent users. Finally, we find a positive and significant association between cannabis use and recent concentration problems at school in the OLS models. However, these results do not hold when individual fixed-effects are controlled for.

Propensity Score Matching

We now turn the results obtained with PSM. Estimation of the propensity score for both any and frequent cannabis use and is displayed in Table 2. We observe that being a female, having grown-up in a nuclear family and wealth are all negatively associated with the onset of any cannabis use. Baseline tobacco and alcohol consumption increases the probability of cannabis initiation. Interestingly, some psychological traits seem to play a protective role, including persistency and positive attitude towards life. Figure A shows the distribution of the propensity scores in both the treatment and control groups and therefore provides an assessment of common support.

Table 3 reports the ATT estimates for each outcome and also includes the critical values of Γ obtained with Rosenbaum bounds. The interpretation of the critical values is discussed below. The

ATTs for absenteeism are positive and significant for both any and frequent cannabis use with kernel and NN (5) matching[3]. Estimates suggest that cannabis users skip on average 0.5 to 1.2 more school days per month than non-users. Results for school difficulties and poor grades are significant across all matching methods but only when we consider frequent cannabis use. Findings for poor grades are of similar magnitude than those obtained with the fixed-effects specifications and suggest that cannabis users have a 12 to 19% higher probability to obtain poor grades.

To assess the robustness of these results to potential selection bias, we turn to the interpretation of the critical values of Γ. These values reflect the minimum amount of selection on unobservables that would produce treatment effects that are no longer statistically significant. For example, in the case of absenteeism, the critical value of Γ equals 1.25 (NN matching, frequent use), meaning that the presence of unobserved characteristics that would make individuals 25% more likely to be in the treatment group would invalidate the results. Overall, even if results for absenteeism, school difficulties and poor grades are consistently significant, the critical values of Γ never exceed 1.7. However, these bounds reflect “worst-case scenarios.” In other words, they do not indicate the presence of selection bias but only tell use how strong the selection bias should be to invalidate our conclusions.

5. Discussion

In this paper, we investigate the impact of cannabis use on short-term educational outcomes among high school students. We exploit a Swiss longitudinal dataset that follows a cohort of adolescents annually starting at the end of middle school and that collects information on educational outcomes, substance use, and on a wide range of individual characteristics. We investigate six different outcomes and are able to control for a rich set of baseline characteristics at both the individual and family level. Results obtained with two distinct empirical strategies consistently show that cannabis

[3] Bootstrap standard errors are used to assess the statistical significance of the ATT estimates.

users skip school more often and are more likely to obtain poor grades than non-users. In addition, the comparison of the results obtained with the OLS and FE specifications underscores the importance of taking unobserved heterogeneity into account in this type of analysis.

This study has, however, several important limitations. First, self-reported measures are used for both cannabis use and educational outcomes. These two groups of variables may be subject to intentional misreporting and results may therefore suffer from attenuation bias. Second, we are not able to assess whether our findings on *Poor grades* are driven by impaired cognitive ability or by reduced attendance. Additional analyses are needed to investigate these potential mechanisms into more details and to define the proper interventions. Finally, the information on cannabis use only informs use on the frequency of use but does not provide insights in the intensity of use, neither on the context in which the product is more often consumed.

In this group of high ability students, we observe strong effects of cannabis use on an indicator of exposure to schooling (i.e., school days skipped) and on an indicator of performance (i.e., grades). These results are in line with previous findings [6, 13] and should be taken into account in the development of future messages on the risks of cannabis use. With the unclear signals sent by a relatively permissive legislation and an increasingly widespread use of this product for medical purposes, adolescents may underestimate the full consequences of cannabis use. Information campaigns and school-based programs should be implemented to increase awareness among adolescents that even occasional use might impair their ability to effectively engage in school and may reduce their overall performance.

References

1. Gmel, G., Kuendig, H., Maffli, E., Notari, L., Wicki, M., Georges, A., Grisel-Staub, E., Müller, M., Dubois-Arber, F., Gervasoni, J.-P., Lucia, S., Jeannin, A., Uchtenhagen, A., Schaub, M.: Monitorage suisse des addictions / Rapport annuel –Données 2011, Berne (2012)

2. Menghrajani, P., Klaue, K., Dubois-Arber, F. and Michaud, P.-A.: Swiss adolescents' and adults' perceptions of cannabis use: a qualitative study. Health Education Research. 20(4), 476–484 (2005)

3. Fontes, M.A., Bolla, K.I., Cunha, P.J., Almeida, P.P., Jungerman, F., Laranjeira, R.R., Bressan, R.A. and Lacerda, A.L.T.: Cannabis use before age 15 and subsequent executive functioning. The British Journal of Psychiatry. 198(6), 442–447 (2011)

4. Gruber, S.A., Sagar, K.A., Dahlgren, M.K., Racine, M. and Lukas, S.E.: Age of onset of marijuana use and executive function. Psychol Addict Behav. 26(3), 496-506 (2012)

5. Meier, M.H., Caspi, A., Ambler, A., Harrington, H., Houts, R., Keefe, R.S., McDonald, K., Ward, A., Poulton, R. and Moffitt, T.E.: Persistent cannabis users show neuropsychological decline from childhood to midlife. Proceedings of the National Academy of Sciences. 109(40), E2657–E2664 (2012)

6. Roebuck, M.C., French, M.T. and Dennis, M.L.: Adolescent marijuana use and school attendance. Economics of Education Review. 23(2), 133–141 (2004)

7. Bray, J.W., Zarkin, G.A., Ringwalt, C. and Qi, J.: The relationship between marijuana initiation and dropping out of high school. Health Economics. 9(1), 9–18 (2000)

8. Horwood, L.J., Fergusson, D.M., Hayatbakhsh, M.R., Najman, J.M., Coffey, C., Patton, G.C., Silins, E. and Hutchinson, D.M.: Cannabis use and educational achievement: Findings from three Australasian cohort studies. Drug and Alcohol Dependence. 110(3), 247–253 (2010)

9. Lynskey, M.T., Coffey, C., Degenhardt, L., Carlin, J.B. and Patton, G.: A longitudinal study of the effects of adolescent cannabis use on high school completion. Addiction. 98(5), 685–692 (2003)

10. Macleod, J., Oakes, R., Copello, A., Crome, I., Egger, M., Hickman, M., Oppenkowski, T., Stokes-Lampard, H. and Davey Smith, G.: Psychological and social sequelae of cannabis and other illicit drug use by young people: a systematic review of longitudinal, general population studies. The Lancet. 363(9421), 1579–1588 (2004)

11. Yamada, T., Kendix, M. and Yamada, T.: The impact of alcohol consumption and marijuana use on high school graduation. Health Economics. 77 (5), 92-96 (1996)

12. McCaffrey, D.F., Pacula, R.L., Han, B. and Ellickson, P.: Marijuana use and high school dropout: the influence of unobservables. Health Economics, 19(11), 1281–1299 (2009)

13. Pacula, R.L., Ringel, J. and Ross, K.E.: Does marijuana use impair human capital formation? Rand Health Working Paper. WE-121 (2003)

14. Register, C.A., Williams, D.R. and Grimes, P.W.: Adolescent Drug Use and Educational Attainment. Education Economics. 9(1), 1–18 (2001)

15. Van Ours, J.C. and Williams, J.: Why parents worry: Initiation into cannabis use by youth and their educational attainment. Journal of Health Economics. 28(1), 132–142 (2009)

16. Rosenbaum, P. R.: Observational Studies. 2nd ed. Springer, New York (2002)

17. French, M.T. and Popovici, I.: That instrument is lousy! In search of agreement when using instrumental variables estimation in substance use research. Health Economics. 20, 127-146 (2011)

18. Duarte, R., Escario, J.J. and Molina, J.A.: Marijuana consumption and school failure among Spanish students. Economics of Education Review. 25, 472-481 (2006)

19. Dehejia, R.H. and Wahba, S.: Propensity score-matching methods for nonexperimental causal studies. Review of Economics and statistics. 84(1), 151–161 (2002)

20. Rosenbaum, P.R. and Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. Biometrika. 70(1), 41–55 (1983)

21. Abadie, A. and Imbens, G.W.: Bias-Corrected Matching Estimators for Average Treatment Effects. Journal of Business & Economic Statistics. 29(1), 1–11 (2011)

22. Rosenbaum, P.R.: Does a dose-response relationship reduce sensitivity to hidden bias? Biostatistics. 4(1), 1–10 (2003)

23. Rosenbaum, P.R.: Sensitivity analysis in observational studies. In: Everitt, B.S., Howell, D.C. (eds.) Encyclopedia of statistics in behavioral science, pp. 1809-1814. John Wiley & Sons, Ltd, Chichester (2005).

24. Ganzeboom, H.B.G., De Graaf, P.M. and Treiman, D.J.: A standard international socio-economic index of occupational status. Social Science Research. 21(1), 1-56 (1992)

Table 1 – Descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
Cannabis use: average over the three waves				
Any use	0.16	0.37	0	1
Frequent use	0.07	0.25	0	1
Outcomes: average over the three waves				
School days skipped (per month)	1.57	2.49	0	30
School difficulties index	1.00	1.08	0	5
Poor grade	0.51	0.49	0	1
Lack of engagement/attention	0.14	0.35	0	1
Lack of engagement/motivation	0.22	0.42	0	1
Concentration problems	0.13	0.33	0	1
Control variables measured in 2000 (PISA)				
Female	0.62	0.49	0	1
More than 100 books at home	0.79	0.41	0	1
Nuclear family	0.84	0.36	0	1
Index of family wealth [a)]	0.14	0.76	-2.31	3.38
Index of family educational resources [b)]	0.47	0.61	-3.42	0.76
Index of family educational support [c)]	0.01	0.90	-1.49	3.35
Mother had higher education	0.29	0.45	0	1
Father had higher education	0.50	0.50	0	1
Number of siblings	2.1	2	0	20
Index of socioeconomic status [d)]	59.5	16	16	90
Language test score [e)]	5.10	0.85	2	9.5
Math test score [e)]	5.02	0.99	1.8	10
Control variables measured in 2001 (TREE, Wave1)				
Persistency	12.81	3.32	4	16
Self-efficacy	12.40	3.42	4	16
Self-esteem	7.70	5.55	-15	16
Positive attitude towards life	24.15	4.44	6	30
Any alcohol use	0.66	0.47	0	1
Any tobacco use	0.25	0.43	0	1

a) The index of family wealth reflects goods and characteristics of the household (dishwasher, student's own room, Internet connection, number of mobile phones, televisions, computers, cars, and number of bathrooms).
b) The index of family educational resources reflects the availability of a dictionary, a quiet place to study, a desk for study, textbooks, and of calculators at home.
c) The index of family educational support reflects the frequency at which family members are involved with the student's schoolwork: mother, father, and siblings.
d) The PISA International Socio-Economic Index of Occupational Status that ranges from 16 to 90 is used as a measure of socio-economic status (Ganzeboom et al., 1992)
e) Language test score reflects student's ability in which the interview was taken. Math test score reflects student's ability in mathematics.

Table 2 – OLS and FE models for both any and frequent cannabis use

Panel A: Effect of any cannabis use	Absenteeism		School difficulty		Lack of attention		Lack of motivation		Poor grades		Concentration problems	
	OLS	*FE*	*OLS*	*FE*	*OLS*	*FE*	*OLS*	*FE*	*OLS*	*FE*	*OLS*	*FE*
Any cannabis (lag)	0.662***	0.626**	0.081	-0.030	0.090***	0.025	0.086**	0.051	0.043	0.090*	0.034	-0.021
	(0.181)	(0.282)	(0.082)	(0.099)	(0.031)	(0.046)	(0.035)	(0.054)	(0.036)	(0.048)	(0.026)	(0.041)
Panel B: Effect of frequent cannabis use												
	OLS	*FE*	*OLS*	*FE*	*OLS*	*FE*	*OLS*	*FE*	*OLS*	*FE*	*OLS*	*FE*
Frequent cannabis (lag)	1.094***	0.590	0.502***	0.124	0.144***	0.133*	0.065	0.068	0.143***	0.228***	0.130***	0.032
	(0.306)	(0.370)	(0.143)	(0.156)	(0.050)	(0.080)	(0.048)	(0.069)	(0.045)	(0.059)	(0.041)	(0.053)
N	1,867		1,977		2,004		2,006		1,997		2,101	

Each coefficient represents a separate regression. Each OLS model controls for gender, family type, family wealth, family educational resources, family educational support, mother's education, father's education, number of siblings, socio-economic status of the parents, language score, math score, each measured in 2000. Models also control for alcohol consumption, tobacco use, persistency, self-efficacy, self-esteem, and positive attitude, each measured at Wave 1 (i.e., in 2001). A Wave dummy is included in each model. Robust standard errors are in parentheses. Full results are displayed in Appendix B. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3 – Logit regressions for both any and frequent cannabis use

	(1) Any use	(1) Frequent use
Female	-0.432**	-1.148***
	(-2.33)	(-4.50)
More than 100 books at home	-0.022	-0.119
	(-0.10)	(-0.42)
Nuclear family	-0.419**	-0.268
	(-2.00)	(-0.96)
Index of family wealth	-0.250**	-0.269*
	(-2.18)	(-1.76)
Index of family educational resources	0.210	0.130
	(1.55)	(0.73)
Index of family educational support	-0.107	0.017
	(-1.05)	(0.12)
Mother had higher education	0.232	0.504*
	(1.16)	(1.90)
Father had higher education	-0.175	-0.465
	(-0.86)	(-1.64)
Number of siblings	0.046**	0.016
	(1.98)	(0.49)
Index of socioeconomic status	0.007	0.015
	(1.04)	(1.62)
Language test score	0.004	-0.169
	(0.03)	(-0.97)
Math test score	-0.170	-0.039
	(-1.59)	(-0.27)
Any alcohol use	0.942***	0.727**
	(4.06)	(2.06)
Any tobacco use	1.565***	1.891***
	(8.69)	(7.26)
Persistency	-0.170***	-0.263***
	(-3.42)	(-3.85)
Self-efficacy	0.169***	0.156**
	(3.06)	(2.14)
Self-esteem	0.0118	-0.026
	(0.56)	(-0.96)
Positive attitude towards life	-0.103***	-0.074*
	(-3.35)	(-1.87)
Constant	0.428	0.428
	(0.43)	(0.32)
N	1,196	1,196

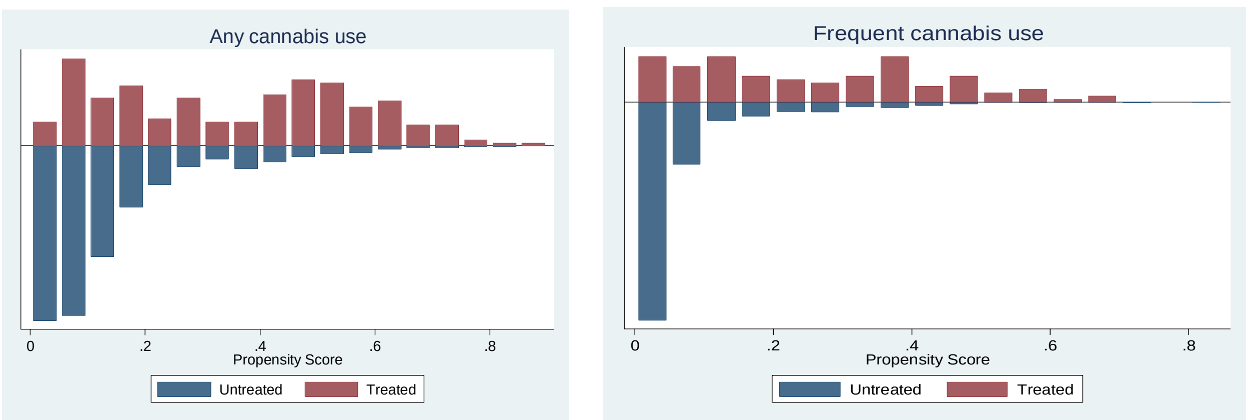
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4 – ATT estimates

	NN	*NN 5*	*Kernel*
Absenteeism			
Any cannabis use	0.440	0.672***	0.594**
Critical level of Gamma	*1*	*1*	*1*
Frequent cannabis use	0.710	1.159***	0.996**
Critical level of Gamma	*1*	*1.35*	*1.15*
School difficulties			
Any cannabis use	-0.086	0.043	0.064
Critical level of Gamma	*1*	*1*	*1*
Frequent cannabis use	0.446*	0.386**	0.370*
Critical level of Gamma	*1.6*	*1.05*	*1.05*
Engagement			
Any cannabis use	0.121*	0.086*	0.093**
Critical level of Gamma	*1.7*	*1*	*1*
Frequent cannabis use	0.067	0.109	0.141***
Critical level of Gamma	*1*	*1*	*1.1*
Motivation			
Any cannabis use	0.096	0.098	0.068
Critical level of Gamma	*1*	*1*	*1*
Frequent cannabis use	0.013	0.019	0.035
Critical level of Gamma	*1*	*1*	*1*
Poor grade			
Any cannabis use	0.048	0.078	0.058
Critical level of Gamma	*1*	*1*	*1*
Frequent cannabis use	0.187**	0.120*	0.146*
Critical level of Gamma	*1.35*	*1.05*	*1.2*
Concentration			
Any cannabis use	0.022	0.001	0.020
Critical level of Gamma	*1*	*1*	*1*
Frequent cannabis use	0.127	0.084	0.083
Critical level of Gamma	*1*	*1*	*1*

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Distribution of the propensity scores



Appendix A - Description of psychological traits variables

Questions framing and response categories

Personality trait	**Framing**	**Possible values**
Persistency	If I decide to accomplish something, I manage to see it through. I complete whatever I start. Even if I encounter difficulties, I persistently continue. I even keep at a painstaking task until I have carried it through.	1 “Completely false” 2 “Mostly false” 3 “Mostly true” 4 “Completely true”
Self-efficacy	I can always manage to solve difficult problems if I try hard enough. I am confident that I could deal efficiently with unexpected events. Thanks to my resourcefulness, I know how to handle unforeseen situations. I can usually handle whatever comes my way.	1 “Completely false” 2 “Mostly false” 3 “Mostly true” 4 “Completely true”
Self-esteem	On the whole, I am satisfied with myself. I feel that I have a number of good qualities. I am able to do things as wells as most of other people. I feel that I am a person of worth, at least on an equal plane with others At times, I think I am not good at all. I certainly feel useless at times. I wish I could have more respect for myself. All in all, I am included to fill that I am a failure.	1 “Not at all true” 2 “Slightly true” 3 “Moderately true” 4 “Very true” 5 “Completely true”
Positive attitude	My future looks bright. I am happy to live. I am happy with the way my life plan unfolds. What ever happens, I can see the positive side of it. My live seems to be meaningful.	1 “Completely false” 2 “Mostly false” 3 “Somewhat false” 4 “Somewhat true” 5 “Mostly true” 6 “Completely true”

In the analyses, each of these psychological traits is measured at baseline (i.e. in Wave 1). For *Persistency*, *Self-efficacy* and *Positive attitude*, we construct indices by simply taking the sum of all items (possible values therefore range from 4 to 16 for *Persistency* and *Self-efficacy* and from 6 to 36 for *Positive attitude*). The index of *Self-esteem* is the sum of the first four items (“positive” self-esteem) minus the sum of the last four items (“negative” self-esteem). *Self-esteem* therefore ranges from -16 to 16.

Appendix B – Full results

Table B1: Any cannabis use – OLS

VARIABLES	(1) Absenteeism	(2) School difficulty	(3) Lack of attention	(4) Lack of motivation	(5) Poor grades	(6) Concentration problems
Any cannabis use (lag)	0.662***	0.0808	0.0896***	0.0857**	0.0430	0.0341
	(0.181)	(0.0822)	(0.0307)	(0.0347)	(0.0358)	(0.0259)
Female	0.130	0.230***	-0.0400*	-0.0445*	-0.0621**	0.0453***
	(0.131)	(0.0601)	(0.0208)	(0.0236)	(0.0280)	(0.0170)
More than 100 books at home	-0.389*	-0.00549	0.00374	-0.0478*	-0.0790**	0.00460
	(0.200)	(0.0757)	(0.0235)	(0.0275)	(0.0330)	(0.0215)
Nuclear family	-0.253	-0.0670	-0.0598**	0.0108	-0.0704**	-0.0147
	(0.194)	(0.0801)	(0.0293)	(0.0282)	(0.0342)	(0.0243)
Index of family wealth	0.198**	-0.0239	0.00977	0.0131	0.0525***	0.0149
	(0.0845)	(0.0395)	(0.0123)	(0.0142)	(0.0180)	(0.0128)
Index of family educational resources	-0.110	-0.0309	-0.0338**	-0.0302*	-0.0125	-0.0178
	(0.100)	(0.0499)	(0.0166)	(0.0177)	(0.0196)	(0.0134)
Index of family educational support	-0.0239	0.0840**	-0.0150	-0.00574	0.000851	-0.0117
	(0.0719)	(0.0337)	(0.0109)	(0.0125)	(0.0141)	(0.00941)
Mother had higher education	0.0792	0.0879	0.00709	0.0581**	-0.00772	-0.00928
	(0.153)	(0.0655)	(0.0215)	(0.0256)	(0.0308)	(0.0193)
Father had higher education	0.0556	-0.0639	-0.0138	-0.0212	-0.0235	0.0141
	(0.145)	(0.0655)	(0.0211)	(0.0239)	(0.0310)	(0.0187)
Number of siblings	-0.00482	-0.00104	0.00000570	0.00570	0.0104***	-0.00209
	(0.0188)	(0.00937)	(0.00321)	(0.00385)	(0.00344)	(0.00203)
Index of socioeconomic status	0.00517	-0.00107	0.000787	0.00133*	-0.00167*	-0.000681
	(0.00444)	(0.00222)	(0.000707)	(0.000728)	(0.000956)	(0.000615)
Language test score	0.0398	0.0413	-0.000417	-0.00138	-0.0192	-0.00825
	(0.0831)	(0.0427)	(0.0137)	(0.0155)	(0.0201)	(0.0121)
Math test score	-0.155**	-0.0728**	-0.00333	0.0241*	-0.0881***	-0.0127
	(0.0703)	(0.0351)	(0.0107)	(0.0131)	(0.0164)	(0.0104)
Any alcohol use	0.155	-0.00977	0.0216	0.0263	-0.0319	-0.00986
	(0.131)	(0.0610)	(0.0183)	(0.0214)	(0.0280)	(0.0167)
Any tobacco use	0.561***	0.121	0.0562**	0.0592**	0.0483	0.0794***
	(0.163)	(0.0759)	(0.0264)	(0.0300)	(0.0330)	(0.0242)
Persistency	-0.0850**	0.0645***	-0.0316***	-0.0286***	-0.0255***	-0.0145***
	(0.0353)	(0.0173)	(0.00592)	(0.00651)	(0.00774)	(0.00503)
Self-efficacy	0.0792**	-0.0912***	0.0122**	0.0151**	-0.00432	0.00337
	(0.0394)	(0.0194)	(0.00615)	(0.00707)	(0.00831)	(0.00569)
Self-esteem	-0.0112	-0.0226***	-0.0000297	0.000916	-0.000389	-0.00736***
	(0.0170)	(0.00840)	(0.00261)	(0.00273)	(0.00322)	(0.00253)
Positive attitude towards life	0.0258	0.00429	-0.00501	-0.0117***	-0.00300	-0.00362
	(0.0231)	(0.0123)	(0.00389)	(0.00416)	(0.00482)	(0.00358)
Wave 3	0.357***	-0.0432	0.00107	0.0163	-0.0147	0.00680
	(0.107)	(0.0363)	(0.0134)	(0.0151)	(0.0168)	(0.0123)
Constant	1.533*	1.495***	0.542***	0.470***	1.764***	0.523***
	(0.805)	(0.376)	(0.123)	(0.135)	(0.152)	(0.112)
N	1,867	1,977	2,004	2,006	1,997	2,101

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

Table B2: Any cannabis use – FE

VARIABLES	(1) Absenteeism	(2) School difficulty	(3) Lack of attention	(4) Lack of motivation	(5) Poor grades	(6) Concentration problems
Any cannabis use (lag)	0.626**	-0.0296	0.0247	0.0511	0.0895*	-0.0211
	(0.282)	(0.0993)	(0.0458)	(0.0542)	(0.0476)	(0.0414)
Wave 3	0.453***	0.000912	0.0167	0.0262*	-0.0130	0.0144
	(0.107)	(0.0347)	(0.0133)	(0.0154)	(0.0171)	(0.0122)
Constant	1.453***	0.998***	0.148***	0.202***	0.518***	0.130***
	(0.0642)	(0.0227)	(0.0101)	(0.0109)	(0.0107)	(0.00873)
N	1,867	1,977	2,004	2,006	1,997	2,101

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

Table B3: Frequent cannabis use – OLS

VARIABLES	(1) Absenteeism	(2) School difficulty	(3) Lack of attention	(4) Lack of motivation	(5) Poor grades	(6) Concentration problems
Frequent cannabis use (lag)	1.094***	0.502***	0.144***	0.0646	0.143***	0.130***
	(0.306)	(0.143)	(0.0503)	(0.0475)	(0.0454)	(0.0412)
Female	0.175	0.255***	-0.0342	-0.0430*	-0.0553*	0.0519***
	(0.132)	(0.0601)	(0.0209)	(0.0236)	(0.0282)	(0.0170)
More than 100 books at home	-0.392**	-0.0104	0.00370	-0.0470*	-0.0796**	0.00408
	(0.200)	(0.0750)	(0.0236)	(0.0276)	(0.0328)	(0.0214)
Nuclear family	-0.250	-0.0600	-0.0598**	0.00936	-0.0692**	-0.0135
	(0.195)	(0.0802)	(0.0296)	(0.0285)	(0.0341)	(0.0239)
Index of family wealth	0.197**	-0.0183	0.00981	0.0120	0.0538***	0.0160
	(0.0847)	(0.0394)	(0.0124)	(0.0143)	(0.0180)	(0.0127)
Index of family educational resources	-0.0997	-0.0274	-0.0325*	-0.0292	-0.0116	-0.0172
	(0.0997)	(0.0489)	(0.0167)	(0.0178)	(0.0196)	(0.0133)
Index of family educational support	-0.0299	0.0835**	-0.0160	-0.00673	0.000469	-0.0120
	(0.0717)	(0.0336)	(0.0109)	(0.0125)	(0.0140)	(0.00936)
Mother had higher education	0.0713	0.0833	0.00537	0.0570**	-0.00940	-0.0102
	(0.153)	(0.0650)	(0.0215)	(0.0257)	(0.0307)	(0.0192)
Father had higher education	0.0685	-0.0555	-0.0122	-0.0211	-0.0213	0.0158
	(0.143)	(0.0650)	(0.0209)	(0.0240)	(0.0309)	(0.0186)
Number of siblings	-0.00756	-0.00383	-0.000394	0.00579	0.00969***	-0.00256
	(0.0185)	(0.00919)	(0.00310)	(0.00383)	(0.00338)	(0.00197)
Index of socioeconomic status	0.00499	-0.00152	0.000760	0.00139*	-0.00177*	-0.000767
	(0.00437)	(0.00222)	(0.000704)	(0.000732)	(0.000955)	(0.000612)
Language test score	0.0485	0.0443	0.00115	-0.000206	-0.0181	-0.00694
	(0.0832)	(0.0427)	(0.0137)	(0.0155)	(0.0200)	(0.0122)
Math test score	-0.164**	-0.0726**	-0.00482	0.0224*	-0.0884***	-0.0133
	(0.0700)	(0.0355)	(0.0108)	(0.0132)	(0.0163)	(0.0104)
Any alcohol use	0.176	-0.0170	0.0245	0.0309	-0.0324	-0.0102
	(0.130)	(0.0609)	(0.0182)	(0.0215)	(0.0279)	(0.0166)
Any tobacco use	0.596***	0.0617	0.0613**	0.0766***	0.0383	0.0697***
	(0.159)	(0.0723)	(0.0262)	(0.0288)	(0.0325)	(0.0230)
Persistency	-0.0831**	0.0700***	-0.0314***	-0.0295***	-0.0244***	-0.0135***
	(0.0351)	(0.0172)	(0.00595)	(0.00653)	(0.00774)	(0.00501)
Self-efficacy	0.0827**	-0.0930***	0.0128**	0.0160**	-0.00453	0.00325
	(0.0395)	(0.0192)	(0.00619)	(0.00712)	(0.00829)	(0.00570)
Self-esteem	-0.00755	-0.0215***	0.000409	0.00120	-0.0000300	-0.00710***
	(0.0169)	(0.00831)	(0.00260)	(0.00274)	(0.00324)	(0.00250)
Positive attitude towards life	0.0221	0.00511	-0.00540	-0.0123***	-0.00294	-0.00352
	(0.0228)	(0.0123)	(0.00387)	(0.00414)	(0.00480)	(0.00352)
Wave 3	0.351***	-0.0528	-0.0000719	0.0168	-0.0170	0.00450
	(0.107)	(0.0363)	(0.0134)	(0.0151)	(0.0168)	(0.0123)
Constant	1.522*	1.415***	0.539***	0.483***	1.747***	0.504***
	(0.790)	(0.377)	(0.122)	(0.136)	(0.153)	(0.110)
N	1,867	1,977	2,004	2,006	1,997	2,101

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

Table B4: Frequent cannabis use - FE

VARIABLES	(1) Absenteeism	(2) School difficulty	(3) Lack of attention	(4) Lack of motivation	(5) Poor grades	(6) Concentration problems
Frequent cannabis use (lag)	0.590	0.124	0.133*	0.0676	0.228***	0.0316
	(0.370)	(0.156)	(0.0798)	(0.0689)	(0.0592)	(0.0534)
Wave 3	0.461***	-0.00368	0.0137	0.0259*	-0.0169	0.0130
	(0.107)	(0.0352)	(0.0134)	(0.0155)	(0.0171)	(0.0123)
Constant	1.512***	0.987***	0.144***	0.206***	0.519***	0.125***
	(0.0539)	(0.0181)	(0.00809)	(0.00804)	(0.00847)	(0.00636)
N	1,867	1,977	2,004	2,006	1,997	2,101

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$