Social Science & Medicine 336 (2023) 116254

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Disentangling the interplay between genes, cognitive skills, and educational level in adolescent and young adult smoking – The TRAILS study

Heiko Schmengler [a,*], Albertine J. Oldehinkel [b], Wilma A.M. Vollebergh [a], Joëlle A. Pasman [c], Catharina A. Hartman [b], Gonneke W.J.M. Stevens [a], Ilja M. Nolte [d], Margot Peeters [a]

[a] Department of Interdisciplinary Social Science, Utrecht University, the Netherlands
[b] Interdisciplinary Center Psychopathology and Emotion Regulation, Department of Psychiatry, University Medical Center of Groningen, University of Groningen, the Netherlands
[c] Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Sweden
[d] Department of Epidemiology, University Medical Center of Groningen, University of Groningen, the Netherlands

ARTICLE INFO

Handling editor: Social Epidemiology Office

Keywords:
Educational level
Smoking
Cognitive skills
Polygenic scores
Adolescence
Young adulthood

ABSTRACT

Recent studies suggest that smoking and lower educational attainment may have genetic influences in common. However, little is known about the mechanisms through which genetics contributes to educational inequalities in adolescent and young adult smoking. Common genetic liabilities may underlie cognitive skills associated with both smoking and education, such as IQ and effortful control, in line with indirect health-related selection explanations. Additionally, by affecting cognitive skills, genes may predict educational trajectories and hereby adolescents' social context, which may be associated with smoking, consistent with social causation explanations. Using data from the Dutch TRAILS Study (N = 1581), we estimated the extent to which polygenic scores (PGSs) for ever smoking regularly (PGS_{SMOK}) and years of education (PGS_{EDU}) predict IQ and effortful control, measured around age 11, and whether these cognitive skills then act as shared predictors of smoking and educational level around age 16, 19, 22, and 26. Second, we assessed if educational level mediated associations between PGSs and smoking. Both PGSs were associated with lower effortful control, and PGS_{EDU} also with lower IQ. Lower IQ and effortful control, in turn, predicted having a lower educational level. However, neither of these cognitive skills were directly associated with smoking behaviour after controlling for covariates and PGSs. This suggests that IQ and effortful control are not shared predictors of smoking and education (i.e., no indirect health-related selection related to cognitive skills). Instead, PGS_{SMOK} and PGS_{EDU}, partly through their associations with lower cognitive skills, predicted selection into a lower educational track, which in turn was associated with more smoking, in line with social causation explanations. Our findings suggest that educational differences in the social context contribute to associations between genetic liabilities and educational inequalities in smoking.

1. Introduction

Lower education has been consistently associated with higher risks of smoking over the life course (Alves et al., 2023). In selective educational systems like in the Netherlands, which are defined by an early selection into different educational tracks based on academic aptitude, educational inequalities in smoking emerge already in early adolescence (de Looze et al., 2013). For example, in 2021, 22.5% of Dutch adolescents (mean age 13.9) in the lower vocational track reported ever experimenting with smoking and 5.5% reported daily smoking, whilst that was only the case for respectively 11.1% and 0.4% of adolescents in the academic track (Boer et al., 2022). Educational trajectories remain strongly associated with smoking after leaving the educational system, and some, but not all, studies even found increases in educational inequalities in smoking between late adolescence and young adulthood (Alves et al., 2023).

Educational inequalities in smoking are thought to emerge in the context of a complex interplay between differences in the social environment and individual differences (e.g., in genetics and cognitive skills), which is not well understood. For example, currently little is known about the mechanisms, including those related to genetically influenced phenotypes, as well as the social context, through which

* Corresponding author.Padualaan 14, 3584, CH, Utrecht, the Netherlands.
E-mail address: h.schmengler@uu.nl (H. Schmengler).

https://doi.org/10.1016/j.socscimed.2023.116254
Received 28 June 2023; Received in revised form 17 August 2023; Accepted 15 September 2023
Available online 20 September 2023

genetic factors may contribute to educational inequalities in smoking in adolescence and young adulthood. This is surprising, given that recent studies have found substantial genetic correlations between smoking and educational attainment (Jang et al., 2022; Quach et al., 2020; Wedow et al., 2018), which suggests that genetic variants associated with smoking and lower educational attainment to some extent overlap. This phenomenon is known as pleiotropy and means that genetic dispositions for smoking are also associated with lower educational attainment, and, similarly, genetic dispositions for lower educational attainment are also associated with smoking, inducing genetic correlations between both phenotypes (Jang et al., 2022). Genetic dispositions for observed outcomes like educational attainment or smoking can be measured with polygenic scores (PGSs). PGSs sum up the effects of a person's many genetic variants on an outcome of interest, or phenotype, using effect sizes estimated in large genome-wide association studies (GWASs) (Allegrini et al., 2022).

Two types of mechanisms may link correlated genetic risk factors for smoking and lower educational attainment to educational inequalities in smoking (i.e., phenotypic correlations between lower education and smoking behaviours). First, genetic dispositions for smoking and lower educational attainment may influence phenotypic characteristics proximally associated with both smoking and lower educational attainment, which can therefore be considered shared risk factors of both. These may include cognitive skills known to be associated with both smoking and educational outcomes, such as IQ (Brody, 1997; Daly and Egan, 2017; Weiser et al., 2010) and effortful control (Daly et al., 2016; deBlois and Kubzansky, 2016; Piehler et al., 2012; Veronneau et al., 2014). The phenomenon that individual characteristics (e.g., genetic risk factors, cognitive skills) related to health behaviours (e.g., smoking) also influence the chance that individuals end up in a certain educational trajectory is known as indirect health-related selection in the public health literature (Mackenbach, 2012). If these characteristics are genetically influenced phenotypes directly associated with both smoking and education (in this case, cognitive skills), this mechanism is also called confounding pleiotropy in the genetics literature (Davies et al., 2019). Alternatively, and also consistent with indirect health-related selection, it is possible that genetic variants affect smoking and education through separate phenotypic mechanisms, which is known as horizontal pleiotropy (Davies et al., 2019).

In a second type of developmental mechanism, like in the first type of mechanism, genes predict into which educational trajectory adolescents are selected, including by means of predicting cognitive skills in childhood. However, unlike in the first type of mechanism, educational differences in the social context then drive associations between genetic risk factors and smoking behaviour. This is referred to as social causation in the public health literature (Mackenbach, 2012). If genetic variants predict one of the two phenotypes of interest (e.g., educational level), which in turn influences the other phenotype of interest (e.g., smoking), this is also referred to as vertical pleiotropy in the genetics literature (Davies et al., 2019). More specifically, adolescents entering the lower educational tracks are much more likely to encounter smoking peers, as classroom social norms in these tracks more strongly encourage smoking (de Looze et al., 2013; Huisman and Bruggeman, 2012; Peeters et al., 2021), making these adolescents more likely to initiate tobacco use themselves. Educational trajectories substantially predict which socio-occupational groups young adults enter (e.g., in terms of occupational class, prestige, and income) (Andersen & Van De Werfhorst, 2010; Behrens et al., 2016; Bol, 2015), and associated differences in social norms, privileges, and stressors, all of which may influence educational differences in smoking behaviour in young adulthood (Huisman et al., 2005; Schaap et al., 2008).

Consistent with both types of mechanisms, associations between PGSs for lower educational attainment and smoking, and between PGSs for smoking and lower educational attainment have been found repeatedly across studies (Hicks et al., 2021; Pasman et al., 2021; Salvatore et al., 2020; Wedow et al., 2018). These associations, which are also referred to as cross-phenotype associations in the genetics literature (Solovieff et al., 2013), may even persist after adjusting for the overlap between PGSs (Hicks et al., 2021). For example, associations between a PGS for having ever smoked regularly and educational attainment, and between a PGS for years of education and smoking remained statistically significant after mutually adjusting for both PGSs (Hicks et al., 2021). This suggests that both PGSs are uniquely related to variance in later smoking and educational attainment and thus should be considered simultaneously when investigating developmental cascades from genetic variants to educational inequalities in smoking.

Research on the developmental mechanisms contributing to the cross-phenotype associations mentioned above is limited. For example, little is known about the processes through which genetic predictors of lower educational attainment are associated with smoking. Only one study among adults found that the association between a PGS for years of education and smoking was partially explained by educational attainment, rather than differences in cognitive ability (Wedow et al., 2018). To our knowledge, this finding has thus far not been replicated in adolescents, which is an important omission given that some research suggests that the contribution of genetics and the environment to smoking behaviour may vary over the course of development (Kendler et al., 2008). Concerning associations between genetic predictors of smoking and educational outcomes, a study among adolescents found that academic motivation, disciplinary problems, and grade point average (GPA) partially mediated the association between a PGS for smoking and educational attainment. These mediating characteristics were also correlated with tobacco use (Hicks et al., 2021). Another study that focussed on associations between PGSs and cognitive skills related to educational attainment (rather than educational attainment directly) found no associations between a PGS for smoking and cognitive ability and executive functioning (Paul et al., 2022). These studies provide mixed evidence on the explanatory mechanisms contributing to cross-phenotype associations between PGSs for smoking and educational attainment. Notably, to our knowledge, no study has thus far explored the interplay between genetic liabilities for smoking, cognitive skills, and educational level in the development of educational inequalities in smoking in adolescents and young adults.

1.1. Aim

In this study, the developmental pathways through which correlated genetic dispositions are associated with educational inequalities in smoking throughout adolescence and young adulthood were investigated. Hereby, this study provides novel insights into the mechanisms underlying the phenotypic associations between educational level and smoking behaviour. We first evaluated whether cross-phenotype associations exist between a PGS for smoking and lower educational attainment, and a PGS for lower educational attainment and smoking, to determine the presence of any form of pleiotropy. Second, we studied the role of indirect health-related selection as developmental pathway linking PGSs to educational differences in smoking. To do so, we evaluated the extent to which both PGSs predict IQ and effortful control measured in childhood (around age 11), and whether these cognitive skills in turn act as shared predictors of both educational level and smoking behaviour in adolescence and young adulthood. We also evaluated whether PGSs serve as shared predictors of smoking and educational level through other mechanisms than IQ and effortful control, which might point at indirect health-related selection via phenotypic mediators we have not measured. Subsequently, we conducted sequentially adjusted regression analyses of associations between educational level and smoking. If indirect health-related selection related to genetic influences and/or cognitive skills is present, these associations should weaken once controlled for PGSs, IQ, and effortful control. Third, we examined the extent to which associations between PGSs and smoking are mediated by the educational trajectories into which adolescents are selected based on their genetic differences, and

hereby educational differences in the social context, consistent with social causation explanations. Our longitudinal approach allows us to consider all phases of adolescent development simultaneously, as the associations between genetic factors, educational level, and smoking may differ across age groups.

2. Methods

2.1. Study population

The TRacking Adolescents' Individual Lives Survey (TRAILS) is a prospective cohort study based in the Netherlands, consisting of a population cohort recruited from primary schools (N = 2229), and a clinical cohort of adolescents recruited from psychiatric outpatient clinics (N = 543), followed from around age 11 onwards. We used data collected during the first six (biennial or triennial) assessment waves, which spanned the period between around age 11 and age 26. The population cohort was invited from 135 schools in the provinces of Groningen, Friesland, and Drenthe, of which 122 decided to participate (de Winter et al., 2005). The clinical cohort consists of adolescents who had been referred to child and adolescent psychiatric outpatient clinics at any point in their life before age 11 for consultation or treatment. The initial response rate was 76% for the general population cohort and 43% for the clinical cohort. A detailed description of TRAILS can be obtained elsewhere (Oldehinkel et al., 2015).

2.2. Genotyping

DNA was available for N = 1694 participants and was collected from blood samples (N = 1334) or, in a minority of participants (N = 360), from buccal swabs (wave 3: general population cohort; wave 2: clinical cohort), and extracted using a manual salting out procedure as discussed in Miller et al. (1988). The Golden Gate Illumina BeadStation 500 and the Infinium™ HumanCytoSNP-12 v2.1 BeadChip platforms (Illumina Inc., San Diego, CA) were used for genotyping. The quality of the genotyping was checked for SNP call rate (>95%), minor allele frequency (>1%), Hardy-Weinberg equilibrium ($p > 1x10^{-6}$), sample call rate (>95%), and heterozygosity (<4SD from mean). Subsequently, datasets were merged, checked for genotype concordance, and imputed using the Haplotype Reference Consortium's global reference panel on the Michigan Imputation Server (Das et al., 2016; McCarthy et al., 2016). Next, we removed participants where at least one parent was born abroad (N = 97). This was done because PGSs based on currently available GWAS, which are mainly based on European-ancestry samples, have inferior prediction accuracy when applied to other ethnic groups (Lee et al., 2018; Mostafavi et al., 2020). Unfortunately, detailed information on participants' ancestry allowing to distinguish between participants with European and non-European migration backgrounds was not available, which is why we decided to exclude all participants with at least one parent born abroad. Lastly, for the small number of sibling pairs in TRAILS, one sibling per pair was randomly removed if genetic data were available for both (N = 16), leading to a final sample of N = 1581 participants.

2.3. Smoking

Adolescents and young adults were asked to report on their tobacco use in the previous four weeks. Responses were recoded to approximate the average number of cigarettes smoked per day. At wave 3 (around age 16), adolescents' answers were coded as follows: 0 (non-smokers), 1 (less than one cigarette per day), 3 (1–5 cigarettes per day), 8 (6–10 cigarettes per day), 15 (11–20 cigarettes per day), 21 (>20 cigarettes per day). From wave 4 to wave 6 (age around 19–26), response options were expanded to capture heavy smoking in more detail: 0 (non-smokers), 1 (less than one cigarette per day), 3 (1–5 cigarettes per day), 8 (6–10 cigarettes per day), 15 (11–20 cigarettes per day), 25 (21–30 cigarettes per day), 31 (>30 cigarettes per day).

2.4. Cognitive skills

Childhood cognitive skills were captured by effortful control and the Intelligence Quotient (IQ) assessed around age 11 (wave 1). Effortful control was assessed using the corresponding subscale from the parent-report Early Adolescent Temperament Questionnaire (EATQ-R), which consists of 11 items with 5 response categories (Cronbach's alpha = 0.86) (Oldehinkel et al., 2004). IQ was estimated using the Block Design and Vocabulary subtests of the Revised Wechsler Intelligence Scale for Children (WISC-R) (Brunnekreef et al., 2007).

2.5. Educational level

The Dutch educational system is characterized by an early (age 11–12) selection into a secondary educational track, based on cognitive tests and the advice of the primary school. There are four tracks in the Dutch educational system, each consisting of a specific type of secondary school followed by tertiary education at the corresponding level (Fig. 1): (1) lower vocational track, (2) intermediate vocational track, (3) higher vocational track, (4) academic track. In addition, there is a special education track, followed by students unable to attend regular education. This track was collapsed with the lower vocational track in our analyses. While in secondary education, adolescents can be recommended by their school to move between educational tracks, depending on their academic performance. Furthermore, after attaining specific milestones of their track, students can become eligible to continue their education in the next higher track. For example, students who finish the intermediate vocational track with an MBO level 4 diploma may continue their education by attending a University of Applied Sciences of the higher vocational track. Overall, a proportion of students was mobile mainly between adjacent educational tracks: 26.70% of adolescents moved to a different track between wave 3 and 4, 26.88% between wave 4 and 5, and 12.51% between wave 5 and 6, respectively. 41.83% of participants were in a different educational track around age 26 (wave 6) than around age 16 (wave 3). Educational track membership was assessed at each wave by asking for participants' current enrolment, as well as their highest completed diploma. Participants who finished the final diploma of a given track received the value corresponding to that level for all subsequent waves, unless they continued education at a higher level.

If information on current and completed education was not available at waves 3 or 4, retrospective event history calendars completed at wave 3 and wave 5 were used to ascertain adolescents' educational level at these waves. It was not possible to classify participants who had not been in education for a longer period, whose educational level was not classifiable in terms of one of the four tracks described above (e.g., because of education abroad), whose educational level was assessed incompletely, who did not respond to questions on education, or who had left the educational system permanently (wave 3: N = 206, 13.32%; wave 4: N = 222, 14.82%, wave 5: N = 240, 16.49%, wave 6: N = 342, 24.95%). Education was considered as missing for these participants.

2.6. Polygenic scores (PGSs)

PGSs for smoking (PGS_{SMOK}) and educational attainment (PGS_{EDU}) were computed as the weighted sum of alleles using LDPred (Vilhjálmsson et al., 2015). Weights (i.e., effect sizes) for PGS_{EDU} were obtained from a large GWAS for years of schooling completed (EduYears; N = 1,131,881; 1271 genome-wide significant loci), based on the total sample with the exception of 23andme (Lee et al., 2018). Weights for PGS_{SMOK} were calculated based on a large GWAS for having ever smoked regularly (SmkInit; N = 1,232,091; 378 associated variants), also based on the total sample with the exception of 23andme (Liu et al., 2019). In prediction analyses, EduYears was able to explain 11–13% of the variance in educational attainment (Lee et al., 2018). SmkInit

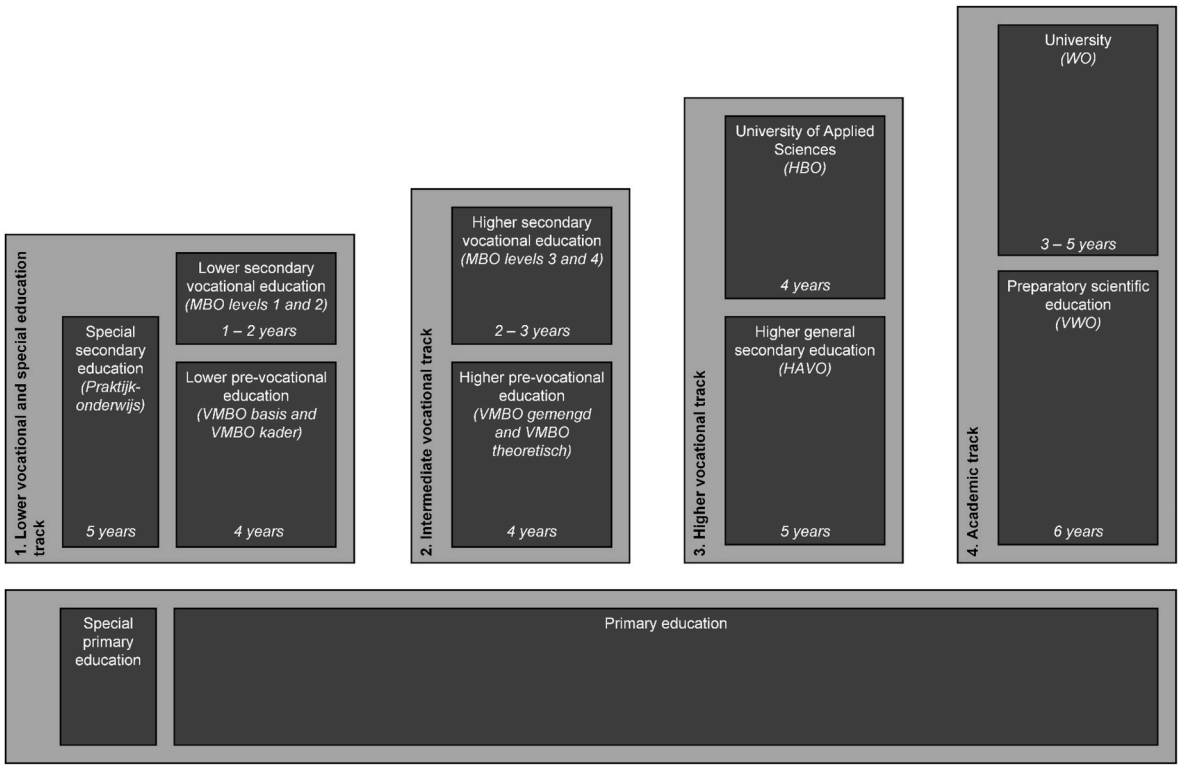

Fig. 1. The Dutch educational system.

accounted for approximately 4% of variance in ever smoking regularly (Liu et al., 2019). The weights were multiplied by the inverse of the linkage disequilibrium scores, as calculated by LDPred from the combined data set of the TRAILS general population and clinical cohort. The most liberal threshold (fraction of causal variants = 1.00) including all SNPs was used, in line with suggestions that this approach best captures the genetic architecture of complex phenotypes, such as education and smoking (Boyle et al., 2017). PGSs were z-score transformed to facilitate interpretability. Higher scores on PGS_{SMOK} represent higher genetic risk for smoking, and higher scores on PGS_{EDU} represent higher genetic risk for lower educational attainment.

2.7. Covariates

Environmentally mediated effects of parental genotypes on offspring phenotypes can induce 'backdoor paths' confounding associations between individual genotypes and phenotypes. This type of confounding, which is also known as dynastic effects (Morris et al., 2020; Pingault et al., 2022), was addressed by controlling for parental educational attainment and smoking. To assess parental educational attainment, the responding parent was asked about their own and their partner's highest educational attainment (wave 1), of which the mean was taken: 1 (elementary education), 2 (lower tracks of secondary education), 3 (higher tracks of secondary education), 4 (higher vocational education), 5 (university). For smoking, the responding parent was asked about their own and their partner's tobacco use in the preceding year (wave 1). Answers were recoded to 0 (neither parent smokes daily), 1 (one parent smokes daily), 2 (both parents smoke daily). We further adjusted for age at baseline, sex, and cohort type (i.e., clinical vs. general population cohort).

2.8. Missing data handling

We performed attrition analyses to evaluate the extent to which dropout may have influenced our findings. Attrition analyses showed that at wave 6 13.28% (N = 210) of our analytic sample no longer participated in the study. Higher scores on both PGSs, lower IQ, lower effortful control, male sex, lower parental education, parental and adolescent smoking, and lower educational level were significantly related to dropout (Table S1). Similar differences were found when comparing participants with classifiable educational level to those whose educational level could not be determined (Table S2). Missing values were addressed using multiple imputations by chained equations under fully conditional specification (van Buuren, 2007) and under the assumption of missingness at random. 90 imputed datasets were created with 50 iterations between datasets.

2.9. Analytical approach

We conducted structural equation models (SEM) in Mplus 8.10 to represent the hypothesized relationships between PGSs, cognitive skills around age 11, and educational level and smoking from around age 16 to 26 (Fig. 2). Separate models were conducted to predict educational level and smoking in each age group (i.e., around age 16, 19, 22, and 26). We first also ran separate models for each PGS, and then combined PGS_{SMOK} and PGS_{EDU} in a single model to account for their overlap and to explore whether each PGS contains variance uniquely associated with educational level and smoking. All regression coefficients were adjusted for all covariates (i.e., parental education and smoking, adolescent age at baseline, sex, and cohort type). We evaluated the potential developmental mechanisms linking PGSs to educational inequalities in smoking by computing total, direct, and indirect effects using the 'model indirect' (mediation) command in Mplus. Besides our SEM models based on Fig. 2, we conducted sequentially adjusted regression models of associations between educational level and smoking around age 16, 19, 22, and 26, to explore the extent to which these associations are explained by individual differences in PGSs, IQ, and effortful control. Smoking was modelled using negative binomial regression to accommodate the zero-inflated nature of this outcome (Allison, 2012). Linear regressions were used to predict IQ, effortful control, and educational level.

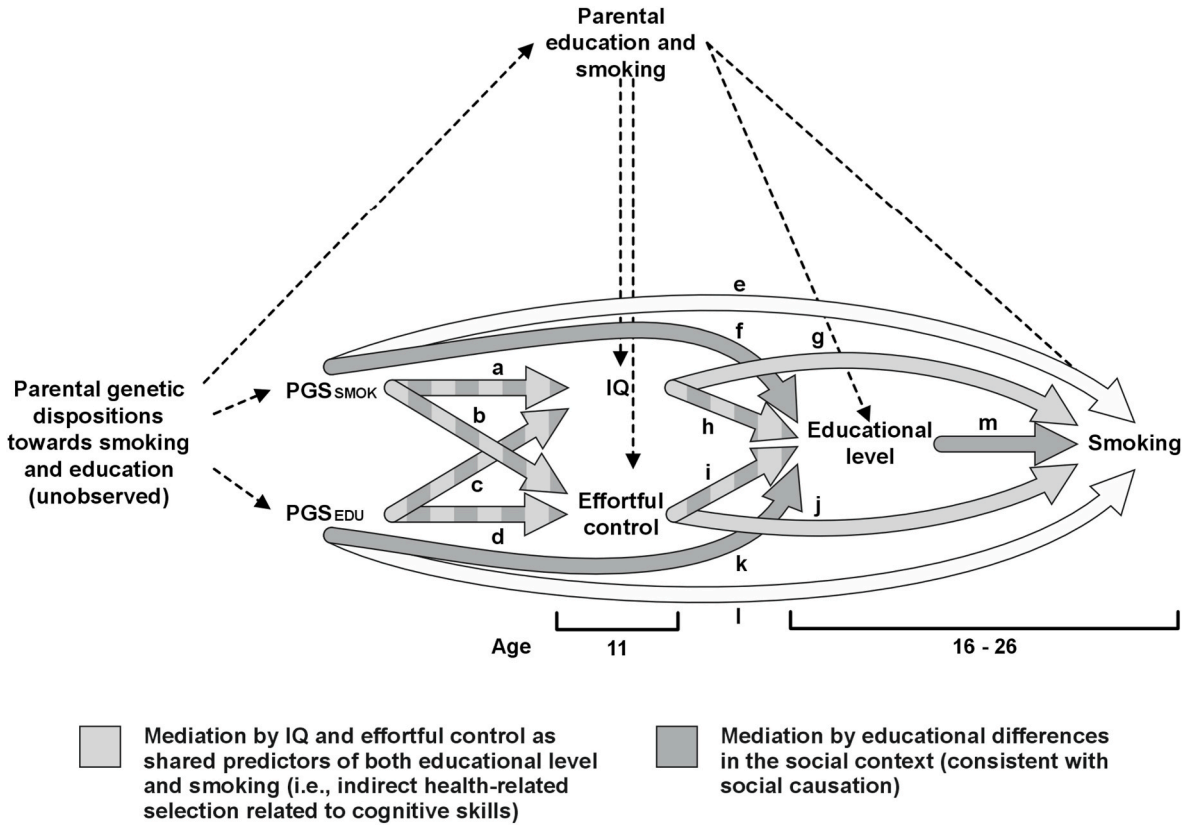

Fig. 2. Conceptual model
Dashed lines represent potential confounding paths. Additional covariates included in all regression equations (omitted from Fig. 2 for clarity) were age at baseline, sex, and cohort type.

Non-normality was accounted for by using robust maximum likelihood (MLR).

We conducted our analyses in three steps:

1.) To determine the existence of cross-phenotype associations (and therefore any form of pleiotropy), we calculated the total effects (i.e., combination of direct and all indirect effects) of PGS_{SMOK} and PGS_{EDU} on both smoking and educational level around age 16, 19, 22, and 26.
2.) To assess the role of indirect health-related selection, we first evaluated whether the PGSs predicted IQ and effortful control at around age 11, and whether these cognitive skills then acted as shared predictors of smoking and educational level measured around ages 16, 19, 22, and 26. We therefore compared the indirect effects of PGS_{SMOK} and PGS_{EDU} on educational level mediated by IQ and effortful control (paths a/h, b/i, c/h, d/i) to the indirect effects of PGS_{SMOK} and PGS_{EDU} on smoking, mediated by IQ and effortful control, but not educational level (paths a/g, b/j, c/g, and d/j). Second, we evaluated direct effects of PGSs on smoking (paths e and l) and educational level (paths f and k). If significant direct effects of a PGS on both smoking and educational level are found, this could, for instance, point at indirect health-related selection through other phenotypic mechanisms than IQ or effortful control. Lastly, we explored potential changes in associations between educational level and smoking (path m) after controlling for PGSs, IQ, and effortful control in sequentially adjusted regression models.
3.) To evaluate the extent to which genes are associated with selection into different educational tracks, and hereby different social contexts, which may then predict smoking behaviour (consistent with social causation), we estimated the indirect effects of PGS_{SMOK} and PGS_{EDU} on smoking from around age 16 to 26 sequentially mediated by IQ or effortful control around age 11 and educational level, assessed at the same age as smoking (i.e., around age 16, 19, 22, and 26) (paths a/h/m, b/i/m, c/h/m, and d/i/m). We also considered indirect effects of PGSs on smoking via educational level which were not also mediated by IQ or effortful control (paths f/m and k/m).

2.10. Sensitivity analyses

We repeated the SEM models with smoking recoded as binary variable capturing daily smoking (yes/no) using the WLSMV estimator and probit regression. These models were also able to treat educational level as ordinal variable. Currently, it is not possible to conduct mediation analysis with categorical mediators combined with outcomes modelled with negative binomial regression in Mplus. In mediation models it is usually preferable to measure mediators and outcomes in consecutive waves. However, the time lags between measurements in TRAILS are rather long (about three years). This means that adolescents are frequently in different social contexts in one wave compared to the next. To adequately assess the consequences of educational level in terms of smoking, a fairly short time interval between measurements is preferable. This is why we modelled educational level and smoking contemporaneously in our main analyses. Nevertheless, we conducted sensitivity analyses in which we allowed for one wave time lag between measurements of educational level and smoking, to assess whether the choice of time lags affected our results. Furthermore, we assessed whether results differed if the analytical sample was restricted to participants of the general population cohort (N = 1248). Lastly, we conducted a complete case analysis to assess how our way of handling missing data affected results.

3. Results

3.1. Descriptive statistics

Descriptive statistics of the study sample are presented in Table 1. Adolescents' tobacco use increased from around age 16 to around age 22, and decreased again from around age 22 to 26. Lower educational level was consistently associated with more smoking in all age groups. The correlation between PGS_{SMOK} and PGS_{EDU} was r = 0.29.

3.2. Structural equation models

3.2.1. Cross-phenotype associations

As expected, both PGS_{SMOK} and PGS_{EDU} were strongly associated with their respective phenotypes, as indicated by significant total effects on smoking and, respectively, educational level in all age groups (Tables 2 and 5). Associations between PGS_{SMOK} and smoking were similar around age 16 and 19, increased slightly by around age 22, and then decreased again by around age 26. Associations between PGS_{EDU} and educational level were similar at all measurement occasions. Significant cross-phenotype associations were also found (Tables 3 and 4). PGS_{EDU} had significant total effects on smoking at all measurement occasions, which decreased in strength as adolescents became young adults and remained significant after controlling for PGS_{SMOK}. Similarly, PGS_{SMOK} had significant total effects on educational level over the entire follow up, which increased between around age 19 and 22. However, the weaker total effects of PGS_{SMOK} on education around age 16 and 19 did not survive adjustment for PGS_{EDU}.

3.2.2. PGSs and cognitive skills as shared predictors of smoking and education

PGS_{SMOK} and PGS_{EDU} significantly predicted lower effortful control (paths b and d), whereas PGS_{EDU}, but not PGS_{SMOK}, also predicted lower IQ (path c) (Figs. 3–4). Both higher IQ and effortful control were, in turn, associated with a higher educational level (paths h and i). Accordingly, we found significant indirect effects of PGS_{EDU} on lower education via both lower IQ and effortful control (paths c/h and d/i), and of PGS_{SMOK} on lower education via lower effortful control only (path b/i) (Tables 4 and 5). However, when considering direct associations between cognitive skills and smoking, we found only one significant direct association between higher effortful control and less smoking around age 22 (path j) when controlling for PGS_{EDU} but not PGS_{SMOK} (Fig. 3B). This association did not lead to any significant indirect effects of PGS_{EDU} on smoking via effortful control but not educational level (path d/j) (Table 3). We found no direct associations of IQ with smoking, and therefore also no significant indirect effects of PGSs on smoking via IQ besides those via educational level (path c/h/m) (Tables 2 and 3).

We found significant direct effects of PGS_{SMOK} and PGS_{EDU} on both smoking and educational level, suggesting that both PGSs serve as shared predictors of smoking and education through mechanisms other than IQ and effortful control. PGS_{EDU} had significant direct effects on educational level in all age groups (Table 5), as well as on smoking around age 16 and 19 (Table 3). PGS_{SMOK} had very large direct effects on smoking (Table 2), and smaller direct effects on educational level in all age groups, be it that the direct effect of PGS_{SMOK} on educational level around age 16 was no longer significant after adjusting for PGS_{EDU} (Table 4). Lastly, sequentially adjusted regression models (Table S3) revealed only very minor changes in associations between lower educational level and smoking, with the largest reduction in the association between educational level and smoking around age 16 after controlling for PGSs. In the fully adjusted models, the association between educational level and smoking increased from about age 16 to 19 and remained stable from age 19 onwards.

3.2.3. PGSs as predictors of educational trajectories, and educational trajectories as predictors of smoking

Large proportions of the associations between PGS_{EDU} and smoking were mediated by being in a lower educational track. The proportions mediated increased from less than 40% around age 16 to around 60–80% in young adulthood (Table 3). PGS_{EDU} was strongly associated with a lower IQ and less effortful control in childhood (paths c and d), which then predicted selection into a lower educational track (paths h and i). PGS_{EDU} also significantly predicted selection into a lower educational track via other, unmeasured mechanisms (path k). Having a lower educational level, in turn, predicted more smoking behaviour (path m). Accordingly, all indirect effects of PGS_{EDU} on smoking via educational level were significant throughout adolescence and young adulthood (i.e., paths c/h/m, d/i/m, and k/m), and remained so after adjusting for PGS_{SMOK} (Table 3).

Mediation by educational level was also present for the associations between PGS_{SMOK} and smoking, but the corresponding indirect effects tended to be smaller than the indirect effects of PGS_{EDU} on smoking via

Table 1
Characteristics of adolescents and young adults in the TRAILS Study.

	Wave 1		Wave 3		Wave 4		Wave 5		Wave 6	
N participants	1581		1547		1498		1455		1371	
Male sex, N (%)	821	(51.93)	798	(51.58)	770	(51.40)	748	(51.41)	690	(50.33)
Age, mean (SD)	11.08	(0.54)	16.15	(0.68)	19.02	(0.60)	22.17	(0.66)	25.70	(0.65)
Parental education, mean (SD)	3.15	(0.88)								
Parental smoking, mean (SD)	0.64	(0.77)								
Clinical cohort, N (%)	333	(21.06)								
IQ, mean (SD)	99.71	(14.67)								
Effortful control, mean (SD)	3.11	(0.72)								
PGS_{EDU}, mean (SD)	0.00	(1.00)								
PGS_{SMOK}, mean (SD)	0.00	(1.00)								
Correlation PGS_{EDU} and PGS_{SMOK}	0.29									
Educational level, N (%)										
Lower vocational & special education			322	(24.01)	162	(12.70)	128	(10.53)	82	(7.97)
Intermediate vocational			372	(27.74)	441	(34.56)	322	(26.50)	257	(24.98)
Higher vocational			303	(22.60)	364	(28.53)	478	(39.34)	398	(38.68)
Academic			344	(25.65)	309	(24.22)	287	(23.62)	292	(28.38)
Smoking, mean (SD)										
All levels			2.30	(5.13)	3.42	(6.55)	3.57	(6.67)	2.65	(5.47)
Lower vocational & special education			4.14	(6.68)	6.57	(9.45)	8.09	(10.21)	6.08	(8.08)
Intermediate vocational			2.35	(5.10)	3.75	(6.42)	4.66	(6.93)	3.02	(5.61)
Higher vocational			1.23	(3.43)	2.01	(4.57)	2.24	(4.79)	1.91	(4.19)
Academic			0.70	(2.73)	1.21	(3.48)	1.18	(3.53)	0.72	(2.67)

SD = standard deviation.

Table 2
Total, direct, and indirect effects of PGS_{SMOK} on smoking around age 16, 19, 22, and 26 in the TRAILS study; potential mediators were measured around age 11 (IQ and effortful control) and concurrently with smoking (educational level); linear and negative binomial regression models (MLR estimator; beta-coefficient, standard error, p-value).

	Smoking around age 16	Smoking around age 19	Smoking around age 22	Smoking around age 26
Model 1				
Total effect	**0.375 (0.075), <0.001**	**0.358 (0.064), <0.001**	**0.423 (0.067), <0.001**	**0.306 (0.069), <0.001**
Direct effect	**0.322 (0.074), <0.001**	**0.296 (0.062), <0.001**	**0.353 (0.065), <0.001**	**0.233 (0.067), <0.001**
Total indirect effect	**0.053 (0.016), 0.001**	**0.062 (0.017), <0.001**	**0.071 (0.016), <0.001**	**0.073 (0.018), <0.001**
Specific indirect effects				
PGS_{SMOK}→effortful control→smoking	0.008 (0.008), 0.345	0.008 (0.006), 0.201	0.009 (0.006), 0.140	0.009 (0.007), 0.165
PGS_{SMOK}→IQ→smoking	−0.001 (0.003), 0.641	−0.001 (0.002), 0.655	0.000 (0.001), 0.960	0.000 (0.002), 0.839
PGS_{SMOK}→education→smoking	**0.030 (0.012), 0.016**	**0.038 (0.013), 0.004**	**0.047 (0.013), <0.001**	**0.049 (0.015), 0.001**
PGS_{SMOK}→effortful control→education→smoking	**0.014 (0.005), 0.004**	**0.014 (0.005), 0.003**	**0.012 (0.004), 0.002**	**0.012 (0.004), 0.003**
PGS_{SMOK}→IQ→education→smoking	0.003 (0.005), 0.543	0.003 (0.006), 0.542	0.003 (0.005), 0.543	0.003 (0.005), 0.543
Model 2				
Total effect	**0.322 (0.075), <0.001**	**0.299 (0.064), <0.001**	**0.384 (0.068), <0.001**	**0.266 (0.069), <0.001**
Direct effect	**0.295 (0.075), <0.001**	**0.265 (0.063), <0.001**	**0.345 (0.066), <0.001**	**0.224 (0.068), 0.001**
Total indirect effect	0.027 (0.015), 0.062	**0.033 (0.015), 0.028**	**0.039 (0.016), 0.013**	**0.042 (0.017), 0.012**
Specific indirect effects				
PGS_{SMOK}→effortful control→smoking	0.005 (0.006), 0.454	0.005 (0.005), 0.284	0.007 (0.005), 0.181	0.007 (0.005), 0.205
PGS_{SMOK}→IQ →smoking	0.004 (0.005), 0.353	0.003 (0.004), 0.364	0.000 (0.003), 0.951	0.001 (0.003), 0.690
PGS_{SMOK}→education→smoking	0.016 (0.011), 0.158	0.023 (0.012), 0.060	**0.029 (0.012), 0.019**	**0.032 (0.014), 0.022**
PGS_{SMOK}→effortful control→education→smoking	**0.010 (0.004), 0.022**	**0.010 (0.004), 0.018**	**0.009 (0.004), 0.017**	**0.009 (0.004), 0.019**
PGS_{SMOK}→IQ→education→smoking	−0.007 (0.005), 0.157	−0.008 (0.005), 0.150	−0.006 (0.004), 0.149	−0.006 (0.004), 0.151

Boldface denotes statistical significance at $p < 0.05$. Models 1 are adjusted for sex, age, cohort type, parental education, and parental smoking; Models 2 are additionally adjusted for PGS_{EDU}.

Table 3
Total, direct, and indirect effects of PGS_{EDU} on smoking around age 16, 19, 22, and 26 in the TRAILS study; potential mediators were measured around age 11 (IQ and effortful control) and concurrently with smoking (educational level); linear and negative binomial regression models (MLR estimator; beta-coefficient, standard error, p-value).

	Smoking around age 16	Smoking around age 19	Smoking around age 22	Smoking around age 26
Model 1				
Total effect	**0.338 (0.076), <0.001**	**0.330 (0.058), <0.001**	**0.255 (0.057), <0.001**	**0.234 (0.066), <0.001**
Direct effect	**0.227 (0.076), 0.003**	**0.214 (0.060), <0.001**	0.112 (0.059), 0.057	0.097 (0.069), 0.158
Total indirect effect	**0.111 (0.026), <0.001**	**0.116 (0.022), <0.001**	**0.142 (0.023), <0.001**	**0.136 (0.027), <0.001**
Specific indirect effects				
PGS_{EDU}→effortful control→smoking	0.008 (0.010), 0.423	0.009 (0.007), 0.194	0.013 (0.007), 0.069	0.011 (0.008), 0.152
PGS_{EDU}→IQ→smoking	−0.031 (0.020), 0.129	−0.023 (0.015), 0.142	−0.003 (0.015), 0.834	−0.013 (0.017), 0.469
PGS_{EDU}→education→smoking	**0.068 (0.018), <0.001**	**0.067 (0.016), <0.001**	**0.081 (0.017), <0.001**	**0.085 (0.020), <0.001**
PGS_{EDU}→effortful control→education→smoking	**0.018 (0.006), 0.002**	**0.017 (0.005), 0.001**	**0.015 (0.004), 0.001**	**0.015 (0.005), 0.001**
PGS_{EDU}→IQ→education→smoking	**0.048 (0.011), <0.001**	**0.045 (0.009), <0.001**	**0.036 (0.007), <0.001**	**0.038 (0.008), <0.001**
Model 2				
Total effect	**0.272 (0.077), <0.001**	**0.262 (0.061), <0.001**	**0.171 (0.059), 0.004**	**0.179 (0.067), 0.007**
Direct effect	**0.172 (0.077), 0.026**	**0.152 (0.062), 0.015**	0.036 (0.061), 0.551	0.050 (0.070), 0.478
Total indirect effect	**0.099 (0.026), <0.001**	**0.110 (0.022), <0.001**	**0.134 (0.024), <0.001**	**0.129 (0.026), <0.001**
Specific indirect effects				
PGS_{EDU}→effortful control→smoking	0.007 (0.009), 0.444	0.007 (0.006), 0.247	0.009 (0.006), 0.148	0.010 (0.007), 0.167
PGS_{EDU}→IQ→smoking	−0.026 (0.021), 0.227	−0.019 (0.016), 0.229	−0.001 (0.016), 0.956	−0.008 (0.018), 0.664
PGS_{EDU}→education→smoking	**0.059 (0.017), <0.001**	**0.061 (0.016), <0.001**	**0.075 (0.017), <0.001**	**0.076 (0.019), <0.001**
PGS_{EDU}→effortful control→education→smoking	**0.014 (0.005), 0.007**	**0.014 (0.005), 0.005**	**0.012 (0.004), 0.004**	**0.012 (0.004), 0.006**
PGS_{EDU}→IQ → education→smoking	**0.046 (0.011), <0.001**	**0.047 (0.010), <0.001**	**0.039 (0.008), <0.001**	**0.039 (0.008), <0.001**

Boldface denotes statistical significance at $p < 0.05$. Models 1 are adjusted for sex, age, cohort type, parental education, and parental smoking; Models 2 are additionally adjusted for PGS_{SMOK}.

educational level, and not always remained significant after adjusting for PGS_{EDU} (Table 2). PGS_{SMOK} predicted lower effortful control (path b), which was subsequently associated with being in a lower educational track (path i). PGS_{SMOK} was also associated with lower education via other mechanisms than IQ or effortful control that we have not measured (path f). Lower educational level in turn predicted increased risks of smoking (path m), as described above. Accordingly, indirect effects of PGS_{SMOK} on smoking via lower effortful control and lower education (path b/i/m), as well as via other unmeasured predictors of lower education were found in all age groups (path f/m) (Table 2). After adjusting for PGS_{EDU}, indirect effects of PGS_{SMOK} via educational level but not effortful control on smoking around age 16 and 19 were no longer significant (path f/m), while the indirect effects via effortful control and educational level remained significant (path b/i/m). Also, the total indirect effect (i.e., indirect effect via IQ, effortful control, and educational level combined) of PGS_{SMOK} on smoking around age 16 was no longer significant after adjusting for overlap with PGS_{EDU}.

3.3. Sensitivity analyses

Analyses with education evaluated as ordinal variable and daily smoking (yes/no) as outcome (Figs. S1–S2, Tables S4–S7), those with smoking measured one wave after educational level (Figs. S3–S4, Tables S8–S9), and those restricted to participants of the general

Table 4
Total, direct, and indirect effects of PGS_{SMOK} on educational level around age 16, 19, 22, and 26 in the TRAILS study; potential mediators were measured around age 11 (IQ and effortful control); linear and negative binomial regression models (MLR estimator; beta-coefficient, standard error, p-value).

	Educational level around age 16	Educational level around age 19	Educational level around age 22	Educational level around age 26
Model 1				
Total effect	**−0.094 (0.026), <0.001**	**−0.091 (0.023), <0.001**	**−0.105 (0.023), <0.001**	**−0.107 (0.024), <0.001**
Direct effect	**−0.060 (0.021), 0.005**	**−0.062 (0.019), 0.001**	**−0.079 (0.020), <0.001**	**−0.082 (0.022), <0.001**
Total indirect effect	**−0.034 (0.015), 0.019**	**−0.029 (0.012), 0.019**	**−0.025 (0.011), 0.017**	**−0.024 (0.011), 0.020**
Specific indirect effects				
PGS_{SMOK}→effortful control→educational level	**−0.028 (0.008), 0.001**	**−0.023 (0.007), 0.001**	**−0.021 (0.006), 0.001**	**−0.020 (0.006), 0.001**
PGS_{SMOK}→IQ→educational level	−0.007 (0.011), 0.538	−0.006 (0.009), 0.539	−0.005 (0.008), 0.539	−0.005 (0.008), 0.539
Model 2				
Total effect	−0.038 (0.026), 0.149	−0.043 (0.023), 0.062	**−0.055 (0.023), 0.017**	**−0.057 (0.024), 0.019**
Direct effect	−0.033 (0.022), 0.133	**−0.039 (0.020), 0.048**	**−0.051 (0.021), 0.014**	**−0.054 (0.022), 0.015**
Total indirect effect	−0.005 (0.014), 0.725	−0.004 (0.012), 0.724	−0.004 (0.010), 0.668	−0.004 (0.010), 0.723
Specific indirect effects				
PGS_{SMOK}→effortful control→educational level	**−0.020 (0.008), 0.010**	**−0.017 (0.007), 0.010**	**−0.015 (0.006), 0.011**	**−0.014 (0.006), 0.011**
PGS_{SMOK}→IQ→educational level	0.016 (0.010), 0.134	0.013 (0.009), 0.135	0.011 (0.007), 0.136	0.011 (0.007), 0.137

Boldface denotes statistical significance at $p < 0.05$. Models 1 are adjusted for sex, age, cohort type, parental education, and parental smoking; Models 2 are additionally adjusted for PGS_{EDU}.

Table 5
Total, direct, and indirect effects of PGS_{EDU} on educational level around age 16, 19, 22, and 26 in the TRAILS study; potential mediators were measured around age 11 (IQ and effortful control); linear and negative binomial regression models (MLR estimator; beta-coefficient, standard error, p-value).

	Educational level around age 16	Educational level around age 19	Educational level around age 22	Educational level around age 26
Model 1				
Total effect	**−0.256 (0.026), <0.001**	**−0.220 (0.023), <0.001**	**−0.230 (0.023), <0.001**	**−0.230 (0.025), <0.001**
Direct effect	**−0.131 (0.022), <0.001**	**−0.114 (0.020), <0.001**	**−0.142 (0.021), <0.001**	**−0.142 (0.023), <0.001**
Total indirect effect	**−0.125 (0.015), <0.001**	**−0.106 (0.013), <0.001**	**−0.089 (0.011), <0.001**	**−0.088 (0.011), <0.001**
Specific indirect effects				
PGS_{EDU}→effortful control→educational level	**−0.034 (0.009), <0.001**	**−0.029 (0.008), <0.001**	**−0.025 (0.007), <0.001**	**−0.024 (0.006), <0.001**
PGS_{EDU}→IQ→educational level	**−0.091 (0.011), <0.001**	**−0.077 (0.010), <0.001**	**−0.063 (0.009), <0.001**	**−0.063 (0.009), <0.001**
Model 2				
Total effect	**−0.246 (0.027), <0.001**	**−0.209 (0.024), <0.001**	**−0.217 (0.024), <0.001**	**−0.215 (0.026), <0.001**
Direct effect	**−0.122 (0.023), <0.001**	**−0.104 (0.021), <0.001**	**−0.129 (0.022), <0.001**	**−0.128 (0.024), <0.001**
Total indirect effect	**−0.124 (0.016), <0.001**	**−0.105 (0.013), <0.001**	**−0.088 (0.012), <0.001**	**−0.087 (0.012), <0.001**
Specific indirect effects				
PGS_{EDU}→effortful control→educational level	**−0.029 (0.009), 0.001**	**−0.024 (0.008), 0.001**	**−0.021 (0.007), 0.001**	**−0.020 (0.006), 0.002**
PGS_{EDU}→IQ→educational level	**−0.095 (0.012), <0.001**	**−0.081 (0.010), 0.010**	**−0.067 (0.009), <0.001**	**−0.067 (0.009), <0.001**

Boldface denotes statistical significance at $p < 0.05$. Models 1 are adjusted for sex, age, cohort type, parental education, and parental smoking; Models 2 are additionally adjusted for PGS_{SMOK}.

population cohort were largely consistent with our main analyses (Figs. S5–S6, Tables S10–S13). We found some significant direct associations between cognitive skills and smoking in models evaluating PGSs separately, which all lost significance once both PGSs were included in the models. Only in the complete case analysis we found a weakly significant direct association ($p = 0.044$) of IQ on smoking around age 16 that survived adjustment for both PGSs, be it that this association did not result in a significant indirect effect of either PGS on smoking. Otherwise, the complete case analyses were mostly consistent with our main results (Figs. S7–S8, Tables S14–S17).

4. Discussion

4.1. Summary of results

We investigated the developmental pathways through which genetic dispositions contribute to educational inequalities in smoking throughout adolescence and young adulthood. Genetic vulnerability for smoking (PGS_{SMOK}) was associated with having a lower educational level throughout adolescence and young adulthood. Similarly, a PGS for having a lower educational attainment (PGS_{EDU}) was associated with smoking in all age groups. Associations between PGS_{SMOK} and lower education strengthened, while associations between PGS_{EDU} and smoking weakened as adolescents became young adults. Most of these associations remained significant after mutually adjusting for both PGSs. Whereas PGS_{SMOK} and PGS_{EDU} were both significantly associated with lower effortful control, and PGS_{EDU} also with lower IQ, direct associations between cognitive skills and smoking were absent once controlled for covariates and both PGSs. This suggests that cognitive skills largely do not serve as shared predictors of educational level and smoking. Accordingly, changes in associations between educational level and smoking after adjusting for IQ and effortful control were negligible, suggesting that the role of indirect health-related selection related to these variables may be minor. Instead, associations between both PGSs and smoking seemed partially driven by educational differences in the social context. Partly through associations with lower cognitive skills, PGS_{EDU} and PGS_{SMOK} predicted selection into a lower educational track, which in turn predicted increased smoking behaviour, consistent with social causation explanations.

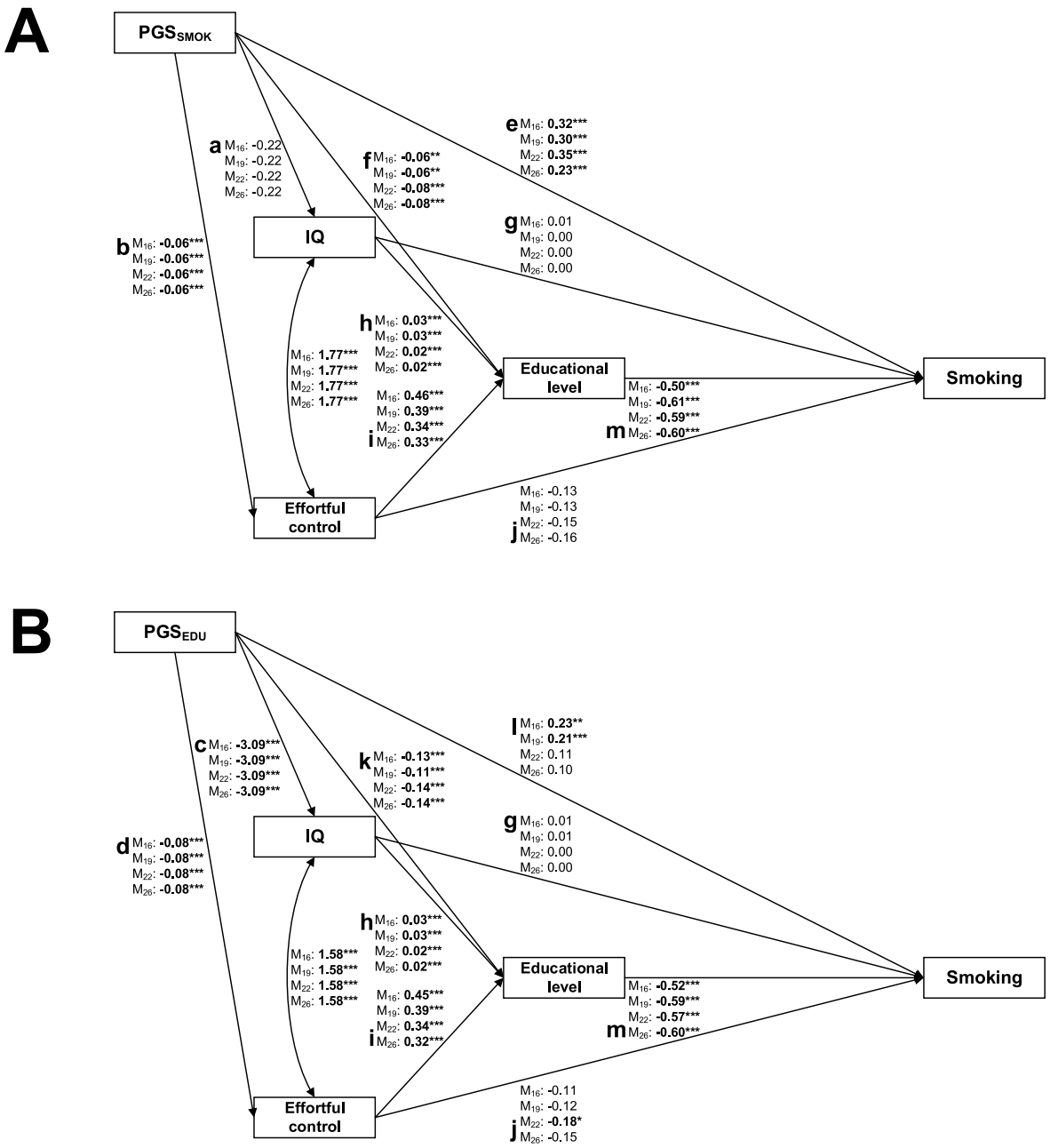

Fig. 3. The associations of PGS_{SMOK} (A) and PGS_{EDU} (B) with smoking around age 16, 19, 22, and 26 in separate models; potential mediators were IQ and effortful control (around age 11) and educational level measured concurrently with smoking; linear and negative binomial regression models (MLR estimator; beta-coefficient).

*p < 0.05; **p < 0.01; ***p < 0.001. All regressions were adjusted for age, sex, cohort type, parental education, and parental smoking. Separate models were used to predict smoking around age 16 (M_{16}), 19 (M_{19}), 22 (M_{22}), and 26 (M_{26}). Educational level was measured concurrently with smoking.

4.2. Interpretation of findings

Most of our results do not support the indirect health-related selection hypothesis in relation to cognitive skills. Whilst both IQ and effortful control were strongly associated with educational level, neither of these cognitive skills showed significant direct associations with smoking behaviour once controlled for covariates and both PGSs. Past studies on direct associations of cognitive skills with smoking have yielded inconsistent results, with some research still finding substantial independent associations after adjusting for educational level (Daly and Egan, 2017; Davies et al., 2017; Sanderson et al., 2019). Differences in time lags between assessments of cognitive skills, educational level, and smoking, as well as in the measures used to capture educational level and cognitive skills may have contributed to this heterogeneity in results. Future research could focus on further disentangling under what conditions measures of cognitive skills remain associated with smoking, net of differences in the social context. It is also possible that other genetically influenced aspects of impulsivity, such as sensation-seeking, are stronger proximal risk factors for smoking than effortful control (Mitchell and Potenza, 2014).

Consistent with the idea that there may be other genetically influenced phenotypic shared predictors of both smoking behaviour and education than IQ and effortful control, we found small reductions in associations between lower education and smoking after controlling for PGSs, as well as some significant direct effects of PGSs simultaneously on smoking and educational level. It is also possible that genetic variants

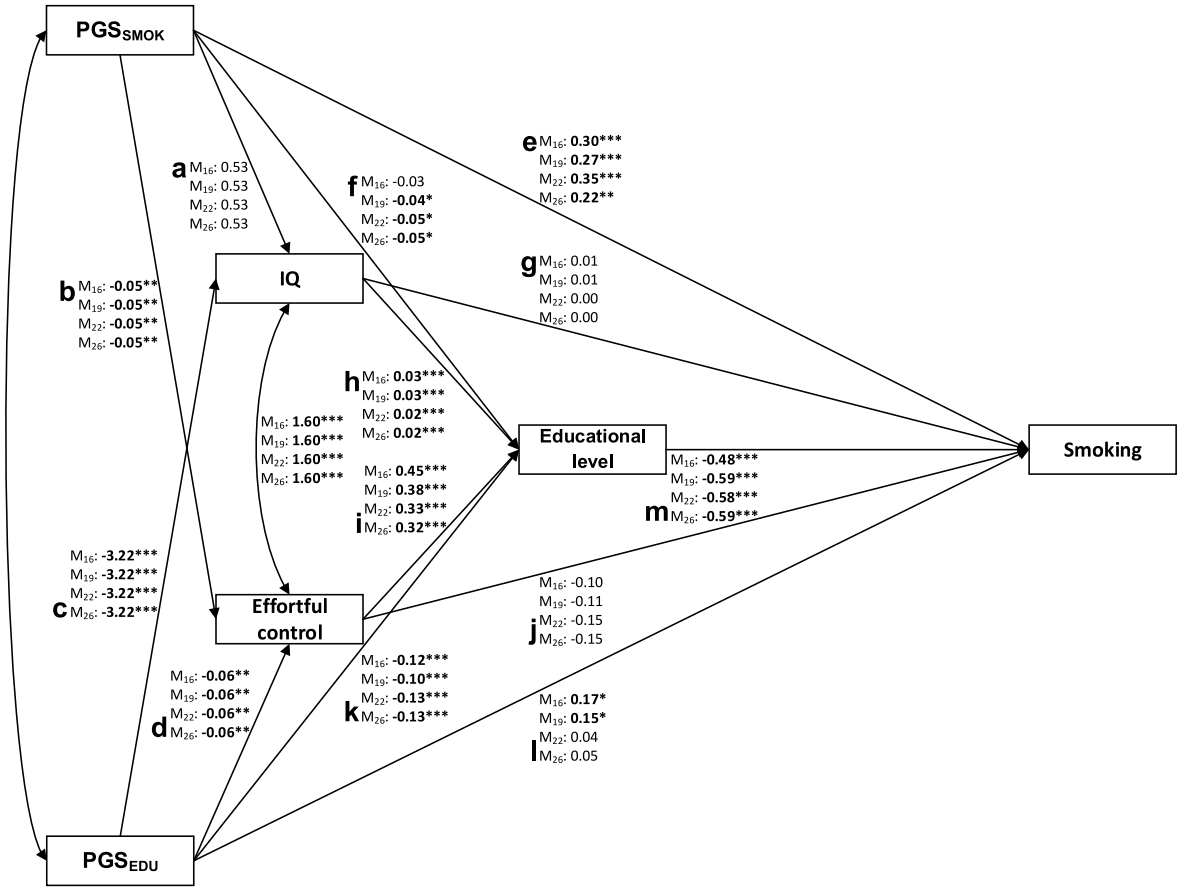

Fig. 4. The associations of PGS_{EDU} and PGS_{SMOK} with smoking around age 16, 19, 22, and 26 in models combining both PGSs; potential mediators were IQ and effortful control (around age 11) and educational level measured concurrently with smoking; linear and negative binomial regression models (MLR estimator; beta-coefficient).

*p < 0.05; **p < 0.01; ***p < 0.001. All regressions were adjusted for age, sex, cohort type, parental education, and parental smoking. Separate models were used to predict smoking around age 16 (M_{16}), 19 (M_{19}), 22 (M_{22}), and 26 (M_{26}). Educational level was measured concurrently with smoking.

influence smoking and educational attainment through separate phenotypic mechanisms (i.e., horizontal pleiotropy) (Davies et al., 2019). Taken together, our results suggests that indirect health-related selection related to unmeasured genetically influenced mediators may still play a role. It is, however, important to note that due to, among other reasons, the limited power of GWAS, currently PGSs capture only part of the genetic variance predictive of smoking and educational attainment (Pingault et al., 2021). Therefore, statistical adjustment for PGSs likely underestimates the contribution of genetics to associations between educational level and smoking.

We found that educational level substantially mediated the associations between PGS_{EDU} and smoking behaviour, and to a lesser extent the associations between PGS_{SMOK} and smoking behaviour. These indirect effects were partly driven by lower cognitive skills in childhood, which predicted selection into a lower educational trajectory. Being in a lower educational trajectory was, in turn, associated with more smoking behaviour, consistent with social causation explanations. Genes, by affecting selection into educational tracks, strongly predict adolescents' options for friendship formation (Huisman and Bruggeman, 2012), exposure to substance use-related social norms, stressors, and perceived future prospects, which may all be related to smoking (Elstad, 2010). Lower educational tracks are more frequently characterized by a culture of futility (Van Houtte and Stevens, 2008) and lower prospects with respect to job/income, potentially leading to increased short-term orientation, and seeking alternative means to attain status amongst peers, which may include risk behaviours like substance use (Elstad, 2010). Correspondingly, classrooms in the vocational educational tracks more strongly feature popularity norms endorsing smoking, and these norms in turn predict adolescents' tobacco use within classrooms (Peeters et al., 2021). The importance of peer effects is further highlighted by social network research demonstrating that the influence of friends on smoking remains strong, even after controlling for friendship selection processes, and that friendship network effects contribute to educational differences in adolescent smoking (Huisman and Bruggeman, 2012). We found that the association between lower education and smoking increased between around age 16 and 19, which is consistent with previous research also showing increases in educational inequalities in smoking towards young adulthood (Alves et al., 2023; Widome et al., 2013). Future research could explore how the transition from education to adult work roles, which often takes place earlier in young adults who followed the vocational tracks (de Looze et al., 2013), may contribute to these increases.

4.3. Strengths and limitations

Key strengths of our study are its high response rate, long follow-up, and consistency of measures over time, allowing to capture multiple developmental periods simultaneously (Oldehinkel et al., 2015). In particular, the selective educational system of the Netherlands provided us a consistent and age-appropriate measure of educational attainment, as proxy for developing socioeconomic status (SES). That is, the selection into educational tracks as early as at age 11–12 years means that Dutch adolescents grow up in distinct educational environments that are characterized by different social norms, future expectations, cognitive

resources, and occupational prospects—characteristics that are closely related to conceptualizations of SES in adulthood. Furthermore, we used very large GWAS for both smoking and educational attainment to calculate PGSs, which (unlike PGSs based on older GWAS) predict similar amounts of variance in phenotypes as many environmental predictors (Lee et al., 2018; Liu et al., 2019). Our longitudinal approach allowed to study differences in associations between genetic propensities, educational level, and smoking behaviours over the course of development, an area that has not been investigated much in past research. Lastly, we were able to address dynastic effects as potential source of confounding by controlling for smoking and educational attainment in parents.

Nevertheless, our study has several limitations. First, attrition and missing data may have affected our results. While we addressed missing data with multiple imputations (van Buuren, 2007), selective missingness in participants with lower educational level, more tobacco use, higher genetic risk for smoking and lower educational attainment, lower parental SES, and lower IQ and effortful control (Tables S1 and S2) may still have influenced our results. Second, to achieve a sufficient sample size, we combined participants from a clinical with a general population cohort, which means that the study sample included more participants with diagnosable psychiatric conditions than would be expected in a representative population-based sample of adolescents. We addressed this issue by including cohort type as covariate in all analyses and by conducting sensitivity analyses that restricted the sample to participants of the general population cohort. Third, we did not control for past levels of smoking in our analyses, as our approach was not focussed on modelling changes in smoking over the course of adolescence, but instead sought to gain insight into the contribution of correlated genetic risk factors to the phenotypic correlations between smoking and educational level that emerge over the course of adolescence. Accordingly, our models could not consider reverse causality in the associations between educational level and smoking. While reverse causation may be less plausible given that nicotine is not intoxicating, smoking could still to some extent be associated with decreases in education, for instance due to long-term effects of nicotine on the developing brain (Yuan et al., 2015).

Fourth, our sample did not include any participants of non-Dutch ethnicity, as large-scale GWAS are currently unavailable for non-European ethnic groups. PGSs based on GWAS from European-ancestry samples tend to have inferior prediction accuracy when applied to other ethnic groups (Lee et al., 2018; Mostafavi et al., 2020), as the frequency of causal alleles and the extent of linkage disequilibrium of SNPs with causal sites differ across populations (Mostafavi et al., 2020). Multiethnic GWAS are necessary to improve external validity, and so is further research on the portability of PGSs across (sub-)populations. Fifth, currently, PGSs capture only part of the genetic variance predictive of smoking and educational attainment. This means that we have only a partial view on the contribution of genetics to the association between adolescent educational level and smoking (Pingault et al., 2021; Wray et al., 2014). At the same time, by providing an individual-level summary measure of the level of genetic risk for a given phenotype, PGSs give a unique opportunity to gain novel insights into the developmental cascades linking correlated genetic risk factors for smoking and lower educational attainment to later educational differences in smoking.

5. Conclusion

Correlated genetic liabilities for smoking and lower educational attainment were significantly associated with both smoking and lower education throughout adolescence and in young adulthood. There was little support for an indirect pathway through cognitive skills (i.e., IQ and effortful control) subsequently acting as shared predictors of educational level and smoking (i.e., no indirect health-related selection related to cognitive skills). Instead, PGSs predicted, partly via their associations with lower cognitive skills, selection into a lower educational trajectory, which in turn predicted more smoking. Our findings shed further light on how social conditions, such as educational differences in the classroom context, add to the genetic relationship between smoking and lower educational attainment. The social contexts in the lower educational tracks (e.g., social norms, peer group composition, social stressors) may therefore be an important target for interventions.

Conflict of interest

The authors declare that they have no conflicts of interest.

Ethical standard statement

Ethical approval for TRAILS was obtained from the Dutch national ethics committee Central Committee on Research Involving Human Subjects (#NL38237.042.11). Written informed consent was obtained from both adolescents (all waves) and their parents (first three waves) prior to each assessment wave.

Author contributions

- Heiko Schmengler: Conceptualization, Methodology, Formal analysis, Visualization, Writing - Original Draft, Writing - Review & Editing.
- Albertine J. Oldehinkel: Supervision, Writing - Review & Editing.
- Wilma A. M. Vollebergh: Conceptualization, Supervision, Project administration, Funding acquisition, Writing - Review & Editing.
- Joëlle A. Pasman: Writing - Review & Editing.
- Catharina A. Hartman: Project administration, Writing - Review & Editing.
- Gonneke W. J. M. Stevens: Writing - Review & Editing.
- Ilja M. Nolte: Formal analysis.
- Margot Peeters: Supervision, Project administration, Writing - Review & Editing.

Data availability

Under the General Data Protection Regulation (GDPR), our dataset is considered pseudonymized rather than anonymized, and is regarded as personal data. When participants were invited to the cohort more than 20 years ago, they were not asked to give informed consent to make their personal data publicly available in pseudonymized form. As a result of this, legal and ethical restrictions prevent the authors from making data from the TRAILS Study publicly available. Data are available upon request from the TRAILS data manager (trails@umcg.nl). Detailed information about the participation agreements with TRAILS participants is available from the ethics committee; Central Committee on Research Involving Human subjects (CCMO; tc@ccmo.nl). For more information about accessing data from the TRAILS Study, please see https://www.trails.nl/en/hoofdmenu/data/data-use. The syntax for our analyses can be obtained from: https://github.com/hschmengler/Interplay-between-genes-cognitive-skills-and-educational-level-in-adolescent-and-young-adult-smoking.

Acknowledgements

This study is made possible by the Consortium on Individual Development (CID). CID is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO) (grant number 024.001.003). This research is part of the TRacking Adolescents' Individual Lives Survey (TRAILS). Organizations participating in TRAILS include various departments of the University Medical Center and University of Groningen, Utrecht University, the Radboud University Medical Center, and the Parnassia Psychiatric Institute, all in the Netherlands. TRAILS has been financially supported by various grants

from the Netherlands Organization for Scientific Research (NWO), ZonMW, GB-MaGW, the Dutch Ministry of Justice (WODC), the European Science Foundation, the European Research Council, BBMRI-NL, and the participating universities. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors would like to thank the participants of the TRAILS Study and all staff involved in the management and execution of this project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2023.116254.

References

Allegrini, A.G., Baldwin, J.R., Barkhuizen, W., Pingault, J.-B., 2022. Research review: a guide to computing and implementing polygenic scores in developmental research. JCPP (J. Child Psychol. Psychiatry) 63 (10), 1111–1124.
Allison, P.D., 2012. Logistic Regression Using SAS - Theory and Application, second ed. SAS Press, Cary, NC, USA.
Alves, J., Perelman, J., Ramos, E., Kunst, A.E., 2023. The emergence of socioeconomic inequalities in smoking during adolescence and early adulthood. BMC Publ. Health 23 (1), 1382.
Andersen, R., Van De Werfhorst, H.G., 2010. Education and occupational status in 14 countries: the role of educational institutions and labour market coordination. Br. J. Sociol. 61 (2), 336–355.
Behrens, T., Groß, I., Siemiatycki, J., Conway, D.I., Olsson, A., Stücker, I., Brüning, T., 2016. Occupational prestige, social mobility and the association with lung cancer in men. BMC Cancer 16 (1), 395.
Boer, M., van Dorsselaer, S., de Looze, M., de Roos, S., Brons, H., van der Einden, R., et al., 2022. HBSC 2021 - Gezondheid en Welzijn van Jongeren in Nederland. Utrecht University, Utrecht, the Netherlands.
Bol, T., 2015. Has education become more positional? Educational expansion and labour market outcomes, 1985–2007. Acta Sociol. 58 (2), 105–120.
Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell 169 (7), 1177–1186.
Brody, N., 1997. Intelligence, schooling, and society. Am. Psychol. 52 (10), 1046–1050.
Brunnekreef, J.A., De Sonneville, L.M.J., Althaus, M., Minderaa, R.B., Oldehinkel, A.J., Verhulst, F.C., Ormel, J., 2007. Information processing profiles of internalizing and externalizing behavior problems: evidence from a population-based sample of preadolescents. JCPP (J. Child Psychol. Psychiatry) 48 (2), 185–193.
Daly, M., Egan, M., 2017. Childhood cognitive ability and smoking initiation, relapse and cessation throughout adulthood: evidence from two British cohort studies. Addiction 112 (4), 651–659.
Daly, M., Egan, M., Quigley, J., Delaney, L., Baumeister, R.F., 2016. Childhood self-control predicts smoking throughout life: evidence from 21,000 cohort study participants. Health Psychol. 35 (11), 1254–1263.
Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Fuchsberger, C., 2016. Next-generation genotype imputation service and methods. Nat. Genet. 48 (10), 1284–1287.
Davies, L.E.M., Kuipers, M.A.G., Junger, M., Kunst, A.E., 2017. The role of self-control and cognitive functioning in educational inequalities in adolescent smoking and binge drinking. BMC Publ. Health 17 (1), 714.
Davies, N.M., Hill, W.D., Anderson, E.L., Sanderson, E., Deary, I.J., Davey Smith, G., 2019. Multivariable two-sample Mendelian randomization estimates of the effects of intelligence and education on health. Elife 8, e43990.
de Looze, M., ter Bogt, T., Hublet, A., Kuntsche, E., Richter, M., Zsiros, E., Vollebergh, W., 2013. Trends in educational differences in adolescent daily smoking across Europe, 2002–10. Eur. J. Publ. Health 23 (5), 846–852.
de Winter, A.F., Oldehinkel, A.J., Veenstra, R., Brunnekreef, J.A., Verhulst, F.C., Ormel, J., 2005. Evaluation of non-response bias in mental health determinants and outcomes in a large sample of pre-adolescents. Eur. J. Epidemiol. 20 (2), 173–181.
deBlois, M.E., Kubzansky, L.D., 2016. Childhood self-regulatory skills predict adolescent smoking behavior. Psychol. Health Med. 21 (2), 138–151.
Elstad, J.I., 2010. Indirect health-related selection or social causation? Interpreting the educational differences in adolescent health behaviours. Soc. Theor. Health 8 (2), 134–150.
Hicks, B.M., Clark, D.A., Deak, J.D., Schaefer, J.D., Liu, M., Jang, S., Vrieze, S.I., 2021. Polygenic scores for smoking and educational attainment have independent influences on academic success and adjustment in adolescence and educational attainment in adulthood. PLoS One 16 (8), e0255348.
Huisman, C., Bruggeman, J., 2012. The social network, socioeconomic background, and school type of adolescent smokers. Int. J. Behav. Dev. 36 (5), 329–337.
Huisman, M., Kunst, A.E., Mackenbach, J.P., 2005. Inequalities in the prevalence of smoking in the European Union: comparing education and income. Prev. Med. 40 (6), 756–764.
Jang, S.-K., Saunders, G., Liu, M., Jiang, Y., Liu, D.J., Vrieze, S., 2022. Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. Psychol. Med. 52 (5), 968–978.
Kendler, K.S., Schmitt, E., Aggen, S.H., Prescott, C.A., 2008. Genetic and environmental influences on alcohol, caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. Arch. Gen. Psychiatr. 65 (6), 674–682.
Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Cesarini, D., 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. 50 (8), 1112–1121.
Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F., Vrieze, S., 2019. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat. Genet. 51 (2), 237–244.
Mackenbach, J.P., 2012. The persistence of health inequalities in modern welfare states: the explanation of a paradox. Soc. Sci. Med. 75 (4), 761–769.
McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Marchini, J., 2016. A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. 48 (10), 1279–1283.
Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16 (3), 1215-1215.
Mitchell, M.R., Potenza, M.N., 2014. Addictions and personality traits: impulsivity and related constructs. Curr Behav Neurosci Rep 1, 1–12.
Morris, T.T., Davies, N.M., Hemani, G., Smith, G.D., 2020. Population phenomena inflate genetic associations of complex social traits. Sci. Adv. 6 (16), eaay0328.
Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J.K., Przeworski, M., 2020. Variable prediction accuracy of polygenic scores within an ancestry group. Elife 9, e48376.
Oldehinkel, A.J., Hartman, C.A., De Winter, A.F., Veenstra, R., Ormel, J., 2004. Temperament profiles associated with internalizing and externalizing problems in preadolescence. Dev. Psychopathol. 16 (2), 421–440.
Oldehinkel, A.J., Rosmalen, J.G.M., Buitelaar, J.K., Hoek, H.W., Ormel, J., Raven, D., et al., 2015. Cohort profile update: the TRacking adolescents' individual Lives Survey (TRAILS). Int. J. Epidemiol. 44 (1), 76-76n.
Pasman, J.A., Demange, P.A., Guloksuz, S., Willemsen, A.H.M., Abdellaoui, A., ten Have, M., Vink, J.M., 2021. Genetic risk for smoking: disentangling interplay between genes and socioeconomic status. Behav. Genet. 52, 92–107.
Paul, S.E., Hatoum, A.S., Barch, D.M., Thompson, W.K., Agrawal, A., Bogdan, R., Johnson, E.C., 2022. Associations between cognition and polygenic liability to substance involvement in middle childhood: results from the ABCD study. Drug Alcohol Depend. 232, 109277.
Peeters, M., Laninga-Wijnen, L., Veenstra, R., 2021. Differences in adolescents' alcohol use and smoking behavior between educational tracks: do popularity norms matter? J. Youth Adolesc. 50 (9), 1884–1895.
Piehler, T.F., Véronneau, M.-H., Dishion, T.J., 2012. Substance use progression from adolescence to early adulthood: effortful control in the context of friendship influence and early-onset use. J. Abnorm. Child Psychol. 40 (7), 1045–1058.
Pingault, J.-B., Allegrini, A.G., Odigie, T., Frach, L., Baldwin, J.R., Rijsdijk, F., Dudbridge, F., 2022. Research review: how to interpret associations between polygenic scores, environmental risks, and phenotypes. JCPP (J. Child Psychol. Psychiatry) 63, 1125–1139.
Pingault, J.-B., Rijsdijk, F., Schoeler, T., Choi, S.W., Selzam, S., Krapohl, E., O'Reilly, P. F., Dudbridge, F., 2021. Genetic sensitivity analysis: adjusting for genetic confounding in epidemiological associations. PLoS Genet. 17 (6), e1009590.
Quach, B.C., Bray, M.J., Gaddis, N.C., Liu, M., Palviainen, T., Minica, C.C., Hancock, D. B., 2020. Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. Nat. Commun. 11 (1), 5562.
Salvatore, J.E., Barr, P.B., Stephenson, M., Aliev, F., Kuo, S.I.-C., Su, J., Dick, D.M., 2020. Sibling comparisons elucidate the associations between educational attainment polygenic scores and alcohol, nicotine and cannabis. Addiction 115 (2), 337–346.
Sanderson, E., Davey Smith, G., Bowden, J., Munafò, M.R., 2019. Mendelian randomisation analysis of the effect of educational attainment and cognitive ability on smoking behaviour. Nat. Commun. 10 (1), 2949.
Schaap, M.M., van Agt, H.M.E., Kunst, A.E., 2008. Identification of socioeconomic groups at increased risk for smoking in European countries: looking beyond educational level. Nicotine Tob. Res. 10 (2), 359–369.
Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M., Smoller, J.W., 2013. Pleiotropy in complex traits: challenges and strategies. Nat. Rev. Genet. 14 (7), 483–495.
van Buuren, S., 2007. Multiple imputation of discrete and continuous data by fully conditional specification. Stat. Methods Med. Res. 16 (3), 219–242.
Van Houtte, M., Stevens, P.A.J., 2008. Sense of futility: the missing link between track position and self-reported school misconduct. Youth Soc. 40 (2), 245–264.
Veronneau, M.H., Racer, K.H., Fosco, G.M., Dishion, T.J., 2014. The contribution of adolescent effortful control to early adult educational attainment. J. Educ. Psychol. 106 (3), 730–743.
Vilhjálmsson, Bjarni J., Yang, J., Finucane, Hilary K., Gusev, A., Lindström, S., Ripke, S., Price, A.L., 2015. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am. J. Hum. Genet. 97 (4), 576–592.
Wedow, R., Zacher, M., Huibregtse, B.M., Mullan Harris, K., Domingue, B.W., Boardman, J.D., 2018. Education, smoking, and cohort change: forwarding a multidimensional theory of the environmental moderation of genetic effects. Am. Socio. Rev. 83 (4), 802–832.
Weiser, M., Zarka, S., Werbeloff, N., Kravitz, E., Lubin, G., 2010. Cognitive test scores in male adolescent cigarette smokers compared to non-smokers: a population-based study. Addiction 105 (2), 358–363.

Widome, R., Wall, M.M., Laska, M.N., Eisenberg, M.E., Neumark-Sztainer, D., 2013. Adolescence to young adulthood: when socioeconomic disparities in substance use emerge. Subst. Use Misuse 48 (14), 1522–1529.

Wray, N.R., Lee, S.H., Mehta, D., Vinkhuyzen, A.A.E., Dudbridge, F., Middeldorp, C.M., 2014. Research review: polygenic methods and their application to psychiatric traits. JCPP (J. Child Psychol. Psychiatry) 55 (10), 1068–1087.

Yuan, M., Cross, S.J., Loughlin, S.E., Leslie, F.M., 2015. Nicotine and the adolescent brain. J. Physiol. 593 (16), 3397–3412.