**A Genetically-Informed Study of the Causal Relationship Between Fertility and Education**

David M. Brazel, Melinda C. Mills, Felix C. Tropf, and Nicola Barban

## Short Abstract

A recent genome-wide association study of age at first sexual intercourse (AFS) and age at first birth (AFB) found genetic correlations (a measure of overlap in the effects of all common genetic variants on two traits) of 0.73 (SE = 0.02) between educational attainment (EA) and AFB and 0.61 (SE = 0.02) between EA and AFS. We use genomic structural equation models to dissect these correlations, examining the underlying causal relationships. The genetic relationship between AFB, AFS, and EA cannot be explained by the genetic components of a wide range of traits, including personality, psychiatric, reproductive, and physical measures, but age at initiation of smoking (AI), a measure of adolescent risk taking, partially mediates the correlation, as does ADHD. In this light, we suggest that the genetics of fertility and education are connected by a common link to adolescent behavioral disinhibition.

## Extended Abstract

An extensive demographic literature documents the existence of a negative correlation between EA and AFB in developed Western countries (Balbo *et al.*, 2013) and shows that the well-known pattern of postponement of childbirth (Figure 1) is a consequence of women's greater opportunities in education and in the workforce. However, AFB is influenced not only by social and environmental forces but also by genetics. Twin studies have found that AFB is heritable but heritability estimates vary by sample, likely reflecting environmental differences and gene-environment interactions (Neiss *et al*., 2002; Tropf *et al*., 2015). The evidence for genetic effects on fertility timing was strengthened further by the publication of a genome-wide association study (GWAS) which discovered ten locations in the genome that are robustly and independently associated with AFB. A prerequisite for fertility is the initiation of sexual behavior. Behavioral genetic studies of adolescent behavior have found that adolescent substance use and abuse is correlated with precocious sexual activity, other "problem" behaviors, and with aggressiveness, impulsivity, and inattention. These correlations seem to be caused by a single latent factor, dubbed behavioral disinhibition, which is highly heritable (Krueger *et al*., 2002; Iacono *et al*., 2008).

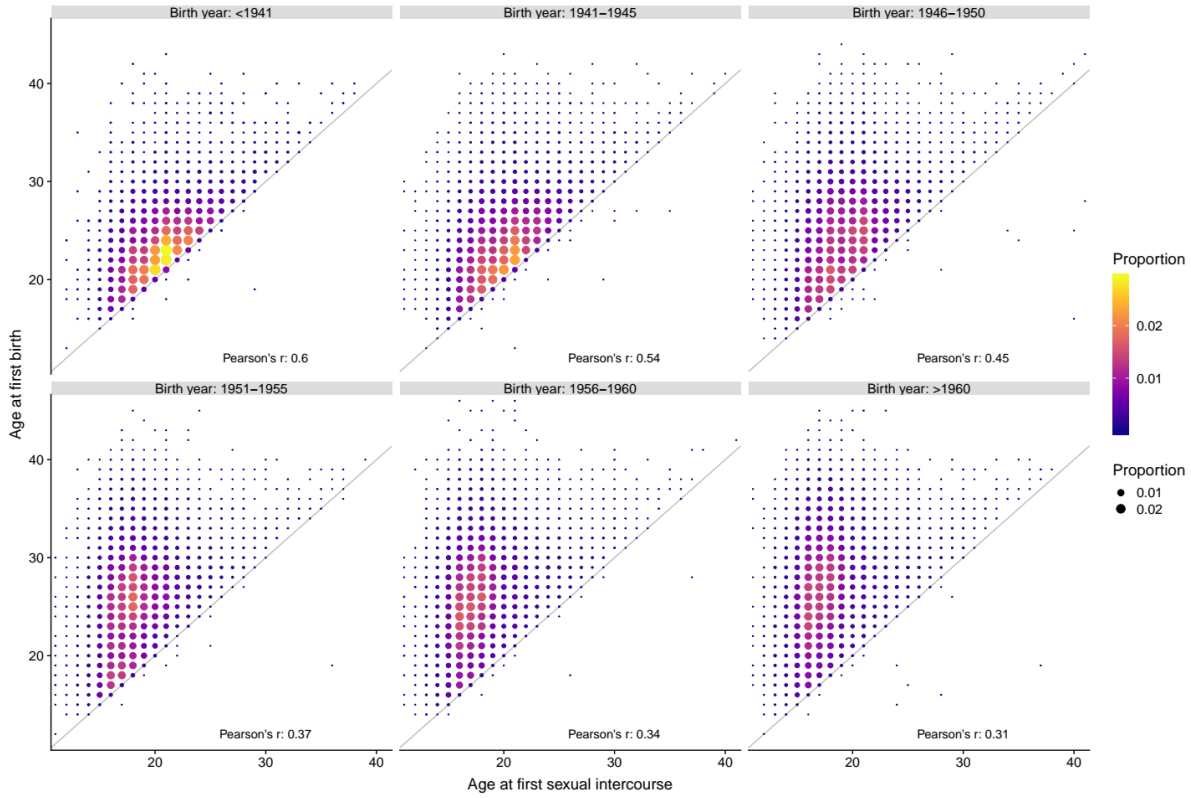


**Figure 1**. A visualization of the relationship between age at first sexual intercourse and age at first birth in the UK Biobank, broken down by birth year.

In the largest GWAS of AFB (N = 542,901) and AFS (N = 387,338) to date, which will be presented separately at EPC 2020 by Prof. Melinda Mills, we identified hundreds of locations across the genome where genetic variation is associated with reproductive behavior. One means of understanding the mechanisms that mediate the genetic effects on a trait is to calculate genetic correlations between that trait and related traits. A genetic correlation is a measure, scaled between 0 and 1, of the overlap in the genetic effects on two traits. A GWAS produces regression weights for millions of genetic variants (mutations) across the genome. LD-score regression (LDSC) is a statistical method that can calculate genetic correlations from these weights (Bulik-Sullivan *et al.*, 2015). We used LDSC to calculate the genetic correlations between AFB, AFS, and a range of putatively associated traits (Figure 2). We find a complex but structured set of correlations. A particularly striking finding is the correlations between AFB and EA *(*rg = 0.73, SE = 0.02) and AFS and EA (rg = 0.61, SE = 0.02).

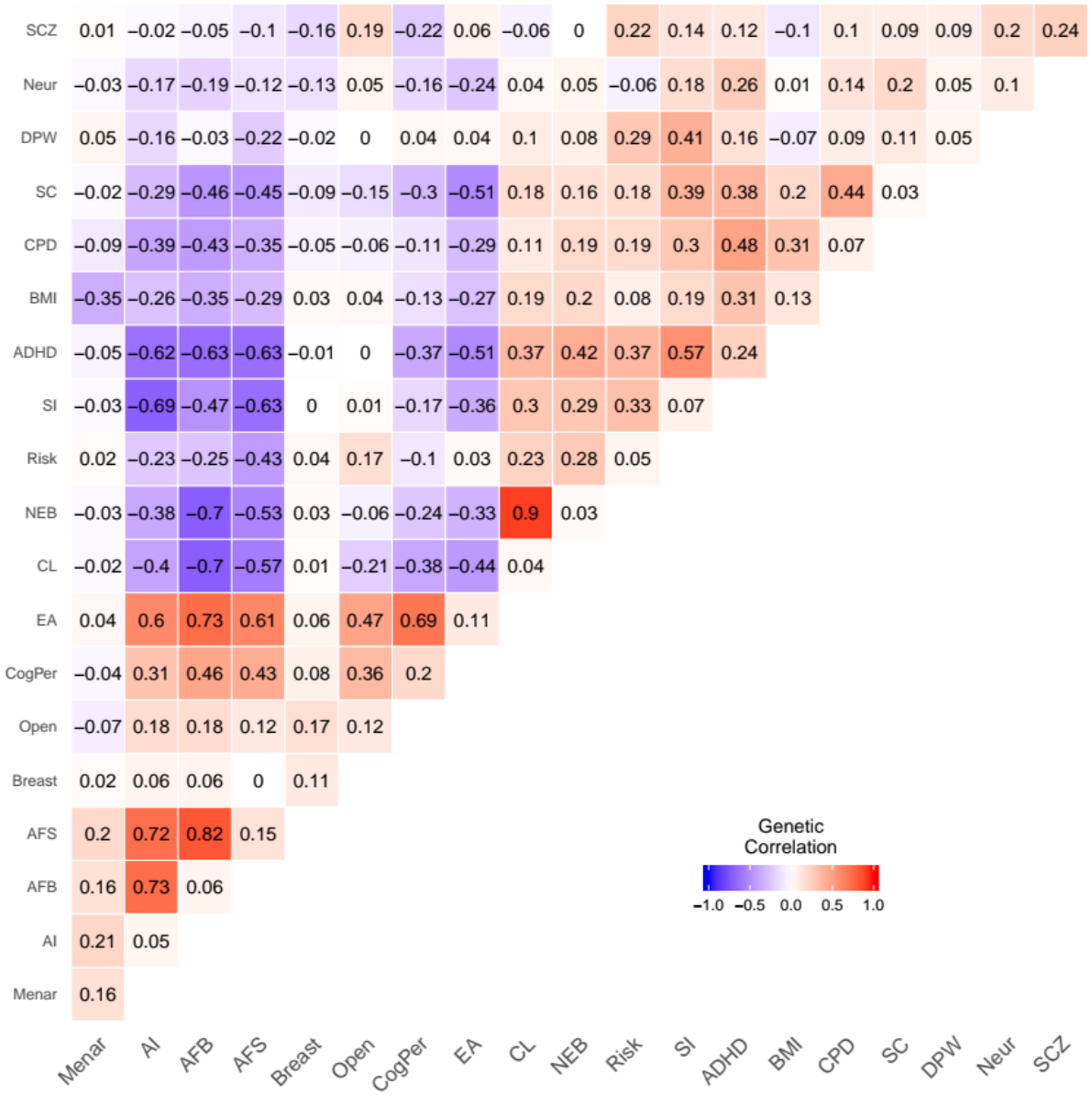


**Figure 2**. A heat map showing the genetic correlations between and among the fertility GWAS phenotypes and relevant phenotypes of interest, as calculated by LD score regression, with SNP heritabilities along the diagonal. Menar = Age at menarche; AI=Age of initiation of smoking; AFB = Age at first birth; AFS = Age at first sex; Breast = Breast cancer; Open = Openness to experience; CogPer = Cognitive performance; EA = Educational attainment; CL = Childlessness; NEB = Number ever born; BMI = Body mass index; CPD = Cigarettes per day; SC = Smoking cessation; Risk = Risk tolerance; SI = Smoking initiation; DPW = Drinks per week; Neur = Neuroticism; SCZ = Schizophrenia; ADHD = Attention deficity hyperactivity disorder

Structural equation models can be used to study the relationship between observed correlations and underlying causal graphs. We used genomic structural equation modeling (GenomicSEM) (Grotzinger *et al*., 2019) to understand the etiology of the correlations between AFB, AFS, and EA by fitting genetic multiple regression models (Figure 3). In a genetic multivariable regression model, some trait C is regressed on traits A and B, producing estimates of the genetic correlation of A with C, independent of B, and of B with C, independent of A. We note that this model is equivalent to a simple mediation model, with C as the dependent variable and either B or C as the mediator. We fit a series of such models in which AFB was regressed on EA and a trait X (Supplementary Tables). In each case, the conditional association of EA and AFB remained

substantial, suggesting that the genetic association of EA with AFB is largely independent of the genetic components of personality (as measured by openness, neuroticism, and risk tolerance), BMI, loneliness, MDD, cognitive performance, substance use, and sexual behaviour (as measured by number of sexual partners). Additionally, the conditional association of cognitive performance with AFB was close to zero, suggesting that cognitive performance does not influence AFB above and beyond its effect on EA. These observations support the conclusion that the genetic correlation between EA and AFB is mediated by environmental mechanisms—those who have high educational attainment have been exposed to an environment that encourages later childbirth. However, we note that the conditional association of EA and AFB was smallest in the model of age of initiation of smoking (AI) and with ADHD. AI is partially genetically distinct from other aspects of cigarette smoking and captures, in part, risk tolerance in adolescence, since most regular smokers initiate in adolescence (Sullivan and Kendler, 1999). AI and ADHD, then, might capture an aspect of adolescent risky behaviour and behavioral disinhibition that our measure of risk tolerance (taken in middle age) does not, explaining its apparent mediation of the relationship between AFB, AFS, and EA. We also fit an analogous series of models in which AFS was regressed on EA (Supplementary Tables), which showed similar patterns of conditional association.

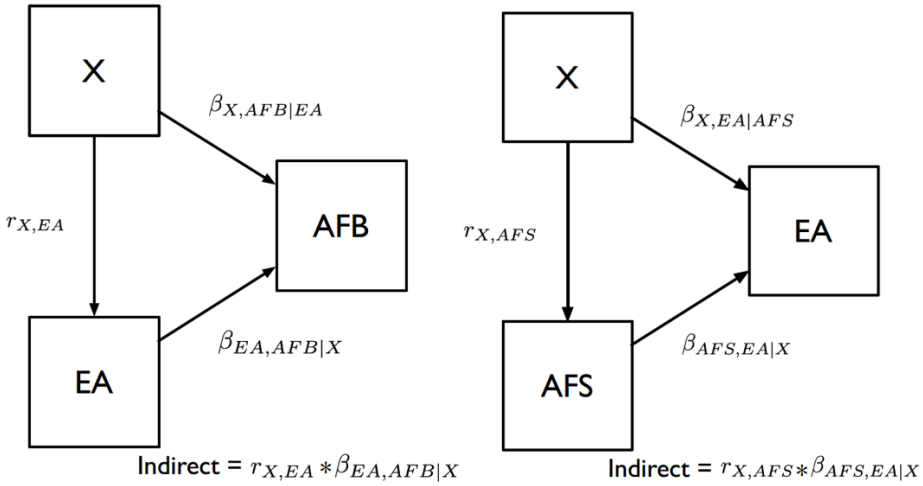


**Figure 3.** Path diagrams of GenomicSEM models showing the genetic multivariable regression model fit to EA and AFB (left) and AFS and EA (right).

We also fit a genetic multivariable regression model in which EA was regressed on AFB and AFS and found a substantial conditional standardized association of EA and AFB (beta = 0.70, SE = 0.05) but a small conditional standardized association of EA and AFS (beta = 0.04, SE = 0.05), as expected since AFS occurs before AFB. We also examined whether the patterns of genetic correlation between AFB and AFS and other traits varied by birth cohort or sex but found little evidence for these forms of heterogeneity. In summary, the genetic components of ADHD and age of initiation of smoking substantially mediate the genetic correlation between AFB, AFS, and EA. This result suggests that behavioral disinhibition plays a significant role not only in adolescent sexual debut but in later-in-life fertility.

**Bibliography**

Balbo, N., Billari, F. C., & Mills, M. (2013). Fertility in Advanced Societies: A Review of Research. *European Journal of Population / Revue Européenne de Démographie*, *29*(1), 1–38. https://doi.org/10.1007/s10680-012-9277-y

Barban, N., Jansen, R., de Vlaming, R., Vaez, A., Mandemakers, J. J., Tropf, F. C., … Mills, M. C. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature Genetics*, *48*(12), 1462–1472. https://doi.org/10.1038/ng.3698

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., … Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211
Grotzinger, A. D., Rhemtulla, M., Vlaming, R. de, Ritchie, S. J., Mallard, T. T., Hill, W. D., … Tucker-Drob, E. M. (2019). Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour*, *3*(5), 513–525. https://doi.org/10.1038/s41562-019-0566-x
Hicks, B. M., Schalet, B. D., Malone, S. M., Iacono, W. G., & McGue, M. (2011). Psychometric and genetic architecture of substance use disorder and behavioral disinhibition measures for gene association studies. *Behavior Genetics*, *41*(4), 459–475. https://doi.org/10.1007/s10519-010-9417-2
Iacono, W. G., Malone, S. M., & McGue, M. (2008). Behavioral Disinhibition and the Development of Early-Onset Addiction: Common and Specific Influences. *Annual Review of Clinical Psychology*, *4*(1), 325–348. https://doi.org/10.1146/annurev.clinpsy.4.022007.141157
Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology*, *111*(3), 411–424. https://doi.org/10.1037/0021-843X.111.3.411
Neiss, M., Rowe, D. C., & Rodgers, J. L. (2002). Does Education Mediate the Relationship Between IQ and Age of First Birth? A Behavioural Genetic Analysis. *Journal of Biosocial Science*, *34*(2), 259–275. https://doi.org/10.1017/S0021932002002596
Sullivan, P. F., & Kendler, K. S. (1999). The genetic epidemiology of smoking. *Nicotine & Tobacco Research*, *1*(785022308), 51–57. https://doi.org/10.1080/14622299050011811
Tropf, F. C., Barban, N., Mills, M. C., Snieder, H., & Mandemakers, J. J. (2015). Genetic influence on age at first birth of female twins born in the UK, 1919–68. *Population Studies*, *69*(2), 129–145. https://doi.org/10.1080/00324728.2015.1056823

## Supplementary Tables

Table A: Standardized results from genetic multivariate regression models examining the relationship between EA and AFB, accounting for the genetic correlation of EA with a third phenotype

| Phenotype X | $r_{EA,AFB}$ | $\beta_{EA,AFB\|X}$ | $r_{X,EA}$ | $r_{X,AFB}$ | $\beta_{X,AFB\|EA}$ | Indirect | StdVar($AFB$) | StdVar($EA$) | StdVar($X$) |
|---|---|---|---|---|---|---|---|---|---|
| Risk tolerance | 0.73 [0.03] | 0.74 [0.02] | 0.03 [0.02] | -0.25 [0.03] | -0.27 [0.02] | 0.02 [0.02] | 0.39 [0.03] | 1.00 [0.03] | 1.00 [0.05] |
| Openness | 0.73 [0.04] | 0.83 [0.06] | 0.47 [0.12] | 0.18 [0.08] | -0.21 [0.10] | 0.39 [0.13] | 0.43 [0.04] | 0.78 [0.07] | 1.00 [0.23] |
| Neuroticism | 0.73 [0.03] | 0.73 [0.02] | -0.24 [0.03] | -0.19 [0.04] | -0.01 [0.04] | -0.17 [0.03] | 0.47 [0.03] | 0.94 [0.03] | 1.00 [0.08] |
| Cognitive performance | 0.73 [0.03] | 0.79 [0.03] | 0.69 [0.01] | 0.46 [0.03] | -0.08 [0.03] | 0.54 [0.07] | 0.46 [0.03] | 0.53 [0.02] | 1.00 [0.03] |
| Age of initiation of smoking | 0.73 [0.03] | 0.46 [0.03] | 0.60 [0.03] | 0.73 [0.06] | 0.46 [0.05] | 0.27 [0.03] | 0.33 [0.03] | 0.64 [0.03] | 1.00 [0.06] |
| Smoking initiation | 0.73 [0.03] | 0.64 [0.02] | -0.36 [0.02] | -0.47 [0.03] | -0.24 [0.02] | -0.23 [0.02] | 0.42 [0.03] | 0.87 [0.03] | 1.00 [0.03] |
| Cigarettes per day | 0.73 [0.03] | 0.66 [0.02] | -0.29 [0.02] | -0.43 [0.04] | -0.24 [0.03] | -0.19 [0.02] | 0.41 [0.03] | 0.91 [0.03] | 1.00 [0.06] |
| Drinks per week | 0.73 [0.03] | 0.73 [0.02] | 0.04 [0.02] | -0.03 [0.02] | -0.06 [0.02] | 0.03 [0.02] | 0.46 [0.03] | 1.00 [0.03] | 1.00 [0.04] |
| Cannabis use | 0.73 [0.03] | 0.72 [0.02] | -0.04 [0.04] | -0.26 [0.07] | -0.23 [0.06] | -0.03 [0.03] | 0.41 [0.03] | 1.00 [0.03] | 1.00 [0.18] |
| Loneliness | 0.73 [0.03] | 0.68 [0.02] | 0.34 [0.03] | 0.37 [0.04] | 0.14 [0.03] | 0.23 [0.03] | 0.45 [0.03] | 0.88 [0.03] | 1.00 [0.06] |
| MDD | 0.73 [0.03] | 0.67 [0.02] | -0.22 [0.03] | -0.41 [0.04] | -0.26 [0.03] | -0.15 [0.02] | 0.40 [0.03] | 0.95 [0.03] | 1.00 [0.06] |

Standard errors of the estimates are provided in brackets. $r_{A,B}$ is the genetic correlation of phenotype A with phenotype B. $\beta_{A,B|C}$ is the standardized regression coefficient relating the genetic component of phenotype A to the genetic component of phenotype B, accounting for the genetic correlation of phenotype A with phenotype C. Indirect is the total indirect effect of X on AFB. StdVar($A$) is the standardized residual genetic variance of phenotype A. EA = Educational attainment; AFB = Age at first birth; MDD = Major depressive disorder.

Table C: Standardized results from genetic multivariate regression models examining the relationship between AFS and EA, accounting for the genetic correlation of AFS with a third phenotype

| Phenotype X | $r_{AFS,EA}$ | $\beta_{AFS,EA\|X}$ | $r_{X,AFS}$ | $r_{X,EA}$ | $\beta_{X,EA\|AFS}$ | Indirect | StdVar($AFS$) | StdVar($EA$) | StdVar($X$) |
|---|---|---|---|---|---|---|---|---|---|
| Risk tolerance | 0.61 [0.03] | 0.76 [0.02] | -0.43 [0.02] | 0.03 [0.02] | 0.36 [0.02] | -0.32 [0.04] | 0.82 [0.03] | 0.52 [0.02] | 1.00 [0.05] |
| Openness | 0.61 [0.03] | 0.56 [0.03] | 0.12 [0.07] | 0.47 [0.12] | 0.40 [0.10] | 0.07 [0.04] | 0.98 [0.04] | 0.46 [0.06] | 1.00 [0.23] |
| Neuroticism | 0.61 [0.03] | 0.59 [0.02] | -0.12 [0.04] | -0.24 [0.03] | -0.17 [0.02] | -0.07 [0.02] | 0.99 [0.03] | 0.59 [0.02] | 1.00 [0.08] |
| Cognitive performance | 0.61 [0.02] | 0.39 [0.02] | 0.43 [0.02] | 0.69 [0.03] | 0.52 [0.01] | 0.17 [0.01] | 0.82 [0.03] | 0.41 [0.02] | 1.00 [0.03] |
| Age of initiation of smoking | 0.61 [0.03] | 0.38 [0.04] | 0.72 [0.04] | 0.60 [0.05] | 0.33 [0.05] | 0.27 [0.05] | 0.49 [0.04] | 0.57 [0.02] | 1.00 [0.06] |
| Smoking initiation | 0.61 [0.03] | 0.64 [0.03] | -0.63 [0.02] | -0.36 [0.02] | 0.04 [0.03] | -0.40 [0.05] | 0.61 [0.03] | 0.62 [0.02] | 1.00 [0.03] |
| Cigarettes per day | 0.61 [0.03] | 0.58 [0.02] | -0.35 [0.03] | -0.29 [0.03] | -0.09 [0.02] | -0.20 [0.02] | 0.88 [0.03] | 0.62 [0.02] | 1.00 [0.06] |
| Drinks per week | 0.61 [0.03] | 0.65 [0.02] | -0.22 [0.02] | 0.04 [0.02] | 0.18 [0.02] | -0.14 [0.02] | 0.95 [0.03] | 0.59 [0.02] | 1.00 [0.04] |
| Cannabis use | 0.61 [0.04] | 0.76 [0.05] | -0.47 [0.09] | -0.04 [0.05] | 0.31 [0.09] | -0.35 [0.10] | 0.78 [0.06] | 0.55 [0.04] | 1.00 [0.18] |
| Loneliness | 0.61 [0.03] | 0.56 [0.02] | 0.35 [0.03] | 0.34 [0.03] | 0.14 [0.02] | 0.20 [0.02] | 0.88 [0.03] | 0.61 [0.02] | 1.00 [0.06] |
| MDD | 0.61 [0.03] | 0.61 [0.02] | -0.37 [0.03] | -0.22 [0.03] | 0.00 [0.03] | -0.22 [0.03] | 0.87 [0.04] | 0.62 [0.02] | 1.00 [0.06] |

Standard errors of the estimates are provided in brackets. $r_{A,B}$ is the genetic correlation of phenotype A with phenotype B. $\beta_{A,B|C}$ is the standardized regression coefficient relating the genetic component of phenotype A to the genetic component of phenotype B, accounting for the genetic correlation of phenotype A with phenotype C. Indirect is the total indirect effect of X on EA. StdVar($A$) is the standardized residual genetic variance of phenotype A. EA = Educational attainment; AFS = Age of first sexual intercourse; MDD = Major depressive disorder.