

Additional file 2: Supplementary Methods

Multiple imputation

We will include all variables in the analytic dataset (14-year questionnaire returners) in the multiple imputation. Age at menarche values will be imputed with a lower bound set to age 14. In addition, to aid the imputation of age at menarche beyond age 14 we will add self-reported growth spurts, body hair development, and skin changes as additional indicators of pubertal stage (using a scale from 0-4, from 'not yet started' to 'already complete'). To validate this approach, we will also set an equivalent percentage of randomly selected individuals to missing, impute their age at menarche and compare the imputed values to their actual values. We will perform multiple imputation using the *mice* package in R, creating 50 datasets. Estimates will then be pooled across the datasets using Rubin's rules. Primary analyses will be conducted on the imputed sample, and results from sensitivity analyses restricted to the complete case sample will be reported in the Supplementary Results of the Stage 2 Registered Report.

Inverse probability weighting

IP weights will be generated based on the results of logistic regression models run in the whole MoBa sample, using return of the 14-year questionnaire participation as the outcome, and all baseline covariates and registry diagnoses (i.e., any variables from the main analytic dataset that do not rely in participation in MoBa beyond the first wave of data collection) as predictors. The models will be run in 50 multiply imputed datasets and derived weights subsequently smoothed by averaging across datasets. Predicted probabilities of participation at 14 years were converted to stabilised weights using the following formula:

$$w(x) = \frac{P(\text{participation} = 1)}{p(x)}$$

Genotyping

Genotyping was performed at three genotyping centres, using different Illumina platforms. The HARVEST project genotyped 33,538 individuals using Illumina HumanCoreExome arrays (12 v.1.1 and 24 v.1.0) at the Genomics Core Facility at the Norwegian University of Science and Technology, Trondheim, Norway. The HARVEST-ERC project genotyped 26,990 individuals using Illumina's Global Screening Array v.1.0 at the Erasmus University Medical Center in Rotterdam, Netherlands. The NORMENT and ADHD/TED projects genotyped 136,275 individuals using Illumina Global Screening Array MD (v.1.0 and v.3.0), 28,977 individuals using Illumina HumanOmniExpress arrays (24v1.0 and 24v1.2), and 9,632 individuals using Illumina Global Screening Array MD v.1.0 + 50k custom OmniExpress overlap content. All NORMENT and ADHD/TED samples were genotyped at deCODE genetics, Reykjavik, Iceland. See further details in Additional file 5.

Psychometric properties

Mothers responded to items on the SMFQ and SCARED using a 3-point scale ranging from "Not true" to "True" to describe their children's symptoms (e.g., depression: "Felt miserable or unhappy", "Thought s/he could never be as good as other kids"; and anxiety respectively: "My child gets really frightened for no reason at all"). In addition, children responded to the SMFQ items at age 14 using the same scale. The original response set for the RS-DBD items was "Never/rarely", "Sometimes", "Often", and "Very often" as mothers rated how often their children engaged in different behaviours. An exception was conduct disorder which adolescents reported at age 14, on a 6-point scale to obtain more variance (information about all measures available in MoBa can be found here: <https://mobawiki.fhi.no/mobawiki/index.php/Questionnaires>). Internal consistency for the SCARED, SMFQ, and RS-DBD subscales at age 8 was generally good to excellent ($N = 20,000$; see Additional files 3 and 7).