Childhood and Adolescence Psychopathology: unravelling the complex etiology by a large Interdisciplinary Collaboration in Europe

D3.1 Poligenic risk Scores

Work Package: WP 3

Lead partner: VU

Author(s): Christel Middeldorp, Meike Bartels & Natascha Stroo

Submission date: May 30, 2019

Version number: 0.1 **Status:** Final

Grant Agreement N°: 721567

Project Acronym: CAPICE

Project Title: Childhood and Adolescence Psychopathology: unravelling the complex etiology by a large Interdisciplinary Collaboration in Europe

Call identifier: H2020-MSCA-ITN-2016

Funding scheme: MSCA-ITN-ETN - European Training Networks

Topic: MSCA-ITN-2016 - Innovative Training Networks

Start date of the project: February 1st, 2017

Duration: 48 months

Dissemination Level	
PU: Public	✓
PP: Restricted to other programme participants (including the Commission)	
RE: Restricted to a group specified by the consortium (including the Commission)	
CO: Confidential, only for members of the consortium (including the Commission)	

MARIE CURIE ACTIONS

European Commission | Horizon 2020 European Union funding for Research & Innovation

Project funded by the European Commission
under the Horizon 2020 Programme
Marie Skłodowska-Curie Actions - Innovative Training Networks

Disclaimer:

Table of Contents

1. Background

Polygenic risk scores (PRS) reflect an individual's genetic vulnerability for a trait. For an overview of the method, we refer to Wray et al (2014). In short, PRS are based on summary statistics from a genome-wide association study performed in the so called discovery sample. In the target sample, the PRS are calculated by adding the number of risk alleles each individual has across the genome, weighted by the effect size of that risk allele. PRS are currently not clinically useful yet, but they can provide important insights for research, for example on whether genetic factors underlie the association between traits. The strength of a PRS analysis depends largely on the size of the discovery sample, and, but to a lesser extent, on the size of the target sample. The tremendous progress in psychiatric genetics provided a great opportunity to calculate polygenic risk scores in the cohorts participating in CAPICE.

2. Discovery samples

PRS were constructed for the adult traits MDD (Wray et al., 2018), schizophrenia (obtained through Psychiatric Genomics Consortium), bipolar disorder (Stahl et al., 2018), educational attainment (Lee et al., 2018), insomnia (Hammerschlag et al., 2017), subjective well-being (Okbay et al., 2016), neuroticism (Okbay et al., 2016), and BMI (Yengo et al., 2018). Height (Yengo et al., 2018) was included as a control phenotype. As the PRS are of higher quality when based on a larger discovery sample, we requested the summary statistics for MDD, bipolar disorder and schizophrenia from the Psychiatric Genomics Consortium, that weren't publicly available at that time. For MDD, cohorts needed to get their own approval because 23andme requires a separate agreement with all parties using their summary statistics.

3. Method to calculate PRS

We limited all analyses to subjects with European ancestry. Genotyping and quality control were performed by each cohort, following common standards. In each cohort, In the case of sample overlap between these GWAS studies and the childhood cohorts of the current study, the childhood cohorts were excluded from the discovery GWAS samples used in the calculation of PRS.
PRS were constructed using LDpred, a method that takes into account the level of linkage disequilibrium (LD) between measured single nucleotide polymorphisms (SNPs) to avoid inflation of effect sizes (Vilhjálmsson et al., 2015). LDpred also allows the inclusion of prior probabilities corresponding to the fraction of SNPs thought to be causal, which allows for testing varying proportions of causal SNPs. As such we tested a range of priors (0.75, 0.5, 0.3, 0.1 and 0.03) in order to assess the prior at which prediction was optimal. We restricted analyses to common variants, using a SNP inclusion criteria of minor allele frequency (MAF) > 5%, and high imputation quality of R2 > 0.9.

The LDpred method and software wasn't always straightforward which led to a delay in the PRS becoming available.
The PRS are now available for all cohorts (see Table 1).

Table 1. Sample characteristics

Cohort	Sample size
ALSPAC	6502
CATSS	11039
GENR	2438
MOBA	4583
NFBC1986	3409
NTR	5501
TEDS	9526
Total	42,998

ALSPAC, Avon Longitudinal Study of Parents and Children; CATSS, Child and Adolescent Twin Study in Sweden; MoBa, Norwegian Mother and Child Cohort Study; NFBC1986, Northern Finland Birth Cohort of 1986; NTR, Netherlands Twin Register; TEDS, Twins Early Development Study