This folder contains the scripts needed to prepare data and run analyses for the paper: "The genetic architecture of differentiating behavioural and emotional problems in early life".

Authors: Adrian Dahl Askelund (jaaskelu@uio.no / ada@lds.no), Laura Hegemann, Andrea G. Allegrini, Elizabeth C. Corfield, Helga Ask, Neil M. Davies, Ole A. Andreassen, Alexandra Havdahl, & Laurie J. Hannigan (laurie.hannigan@bristol.ac.uk).

00\_create\_polygenic\_scores.R creates polygenic scores (PGS) for mothers, fathers, and children, which are merged with the phenotypic data in the script ‘01\_data\_preparation.R’.

00.1\_create\_mdd\_pgs.R creates a new MDD PGS.

01\_data\_preparation.R curates and selects the analytic dataset based on questionnaire, covariate, genotype, and registry data, including data manipulation and preparation of variables. It then creates an illustration of how differentiation is operationalised based on a subset of the raw data.

02\_wave\_specific\_GWAS.R conducts wave-specific GWAS of the difference and total scores in Regenie at the ages of 1.5, 3, and 5 years. Additional preparation steps are conducted by the scripts 02.1-02.3.

02.1\_prepare\_sumstats.R prepares the summary statistics for multivariate GWAS in genomic SEM.

02.2\_munge\_sumstats.R munges the summary statistics before running ld-score regression.

02.3\_ldsc.R runs LDSC on the summary statistics for running multivariate GWAS.

02.4\_run\_multivariate\_GWAS\_diff.R runs multivariate GWAS of the differentiation slope and intercept from an LGM that is equivalent to the phenotypic LGM, but specified based on a genetic variance/covariance matrix.

02.5\_run\_multivariate\_GWAS\_tot.R runs multivariate GWAS of the total problem slope and intercept from the same model as in 02.4, with differentiation and total constrained to be orthogonal (matching their relationship at the phenotypic level).

02.6\_predicted\_sample\_size\_GWAS.R checks the output from the multivariate GWAS and calculates the predicted sample size for the intercept and slope factors.

02.7\_plot\_results\_GWAS.R creates miami plots, manhattan plots, and QQ plots based on the multivariate GWAS output.

03\_genetic\_correlations.R simply states that genetic correlations of differentiation and total problems with the 11 psychiatric and neurodevelopmental conditions are estimated outside of R, in LDSC.

03.1\_calculate\_Neff\_sumstats.R calculates the sample size (N) as the sum of effective sample sizes for each of the 11 GWAS meta-analyses.

03.2\_prepare\_sumstats.R prepares the summary statistics for the 11 psychiatric conditions for genomic SEM.

03.3\_run\_ldsc\_wpsych.R runs ld-score regression of differentiation and total problems, and the 11 psychiatric conditions.

03.4\_plot\_genetic\_correlations.R plots genetic correlations of the 11 psychiatric and neurodevelopmental conditions with differentiation and total problems.

04\_run\_structural\_models.R runs the correlated factor, hierarchical, and bifactor models with factors predicting differentiation and total, by sourcing the script '04.1\_specify\_structural\_models.R'.

04.1\_specify\_structural\_models.R specifies a basic LGM and three (previously established) structural models accounting for variance across 11 psychiatric conditions, which are run by the script '04\_run\_structural\_models.R'.

04.2\_plot\_structural\_models.R creates plots of the structural models of liability to 11 psychiatric conditions, and resulting path estimates.

05\_run\_combined\_LGMs\_with\_pgs.R runs the LGMs with PGS by sourcing the script ‘05.1\_specify\_combined\_LGMs\_with\_pgs.R’.

05.1\_specify\_combined\_LGMs\_with\_pgs.R specifies LGMs with 11 PGS as predictors.

06\_run\_trio\_pgs\_analyses.R runs the LGMs with trio PGS by sourcing the script ‘06.1\_specify\_LGMs\_with\_trio\_pgs.R’.

06.1\_specify\_LGMs\_with\_trio\_pgs.R specifies LGMs with trio PGS as predictors.

07\_run\_mediation\_LGMs\_with\_sx.R runs the mediation LGMs w/ PGS and symptom outcomes by sourcing the script '07.1\_specify\_mediation\_LGMs\_with\_sx.R'.

07.1\_specify\_mediation\_LGMs\_with\_sx.R specifies the mediation LGMs with symptom outcomes.

07.2\_run\_mediation\_LGMs\_with\_dx.R runs mediation LGMs with diagnostic outcomes in Mplus, using the MplusAutomation package in R.

08\_run\_equiv\_tests.R runs equivalence tests to determine, in conjunction with null hypothesis significance tests, whether the PGS effects are null, non-null, or undecided.