Concept Paper Template

Provisional Paper Title: GWAS Meta-Analysis of Aggressive behavior and Attention Problems

Proposing Author: Terrie Moffitt (& Dorret Boomsma)

Author's Email: tem11@duke.edu

Proposed co-authors:
From Duke University:
Karen Sugden, ks164@duke.edu
Avshalom Caspi, ac115@duke.edu

From the Dunedin Study at Univ of Otago:
Richie Poulton Richie.poulton@otago.ac.nz
From the E-Risk Study at KCL:
Louise Arseneault louise.arseneault@kcl.ac.uk

From Free University Amsterdam:
Michel Nivard, m.g.nivard@vu.nl
Meike Bartels, m.bartels@vu.nl
Dorret Boomsma, di.boomsma@vu.nl

P.I. Sponsor:
(if the proposing author is a student or colleague of an original PI)

Today's Date: 12 April 2017

Objective of the study:

The Dunedin and E-Risk studies are being invited to take part as two of many cohorts in a multi-cohort GWAS (genome wide association study). The goal of this initiative is to carry out a meta-analysis of GWAS on Aggressive behavior (AGG) and Attention problems (AP) in subjects between the ages of 3 and 18 years. To attain this goal, phenotype data collected at different ages and rated by different raters (paternal, maternal, self and / or teacher) will be included in a single GWAS meta-analysis, taking dependencies between raters and ages into account. The inclusion of multiple phenotype measures obtained on the same individuals, possibly including subjects rated by more than one rater and more than one age, is somewhat more involved than for a single phenotype. However, for the participating cohorts the analyses consist of running a (possibly large) series of univariate analyses, while the inclusion of repeated measures and measures based on multiple informants will be done at the Meta Analysis level, thereby

substantially increasing power, while still allowing for the expression of age-dependent or rater-dependent genetic effects.

This initial GWAS will be carried out on the E-Risk study using multi-rater aggression, conduct problems and attention phenotypes across ages 5,7,10,12 and 18 (see variables section). This will be followed up by GWAS in the Dunedin Study using multi-rater Antisocial behavior, Aggression and attention phenotypes across ages 5,7,9,11,13,15 and 18. The initial GWAS analysis will be performed at Duke University, genomic data will not be sent elsewhere.

The approach for the cohorts is as follows:
Every cohort will run a separate (univariate) analysis for each phenotype (i.e. one analysis for every combination of rater, age bin and survey measure). On the meta-analysis level, the test statistic per SNP will be combined over rater and age bin, accounting for the dependence between measurements (e.g. the same children rated by father and mother or at multiple ages).

As the meta-analysis will take the form of a multivariate meta regression (see Nivard et al. 2017) it is of the utmost importance to accurately report the covariance between the AGG and AP phenotypes *within* cohort as well as the sample size and sample size overlap as detailed in the SOP. Integrating repeated measures and multiple raters will greatly benefit power. Take, for example, the Netherlands Twin Register dataset where up to ~2700 children are phenotyped or genotyped at 7 ages (3, 7, 10, 12, 14, 16 and 18 years) by either their mother, father or a teacher, this would yield 2700 independent observations to be analyzed in a univariate meta-analysis, while 22.000+ observations are available.

By performing univariate analyses and combining all results in meta-analysis, a loss of data is avoided, and an increase in power is realized. Moreover, this approach allows for a formal test of rater, instrument / diagnosis, age and cohort interaction effects.

These meta-analysis models allow us to not only account for age, rater or instrument effects while increasing power, but also to interpret and disseminate these age and rater specific effects to the broader field of psychiatric genetics. It is of clinical and translational interest to know whether genome-wide hits for adult neurological or psychiatric traits have an effect in childhood. If such an effect is present it is of further interest whether it increases with age or is present from early childhood onward. The results of our developmentally sensitive GWAS could thus be used to enrich and enhance GWAS analysis of adult phenotypes.

Our design allows us to identify to what extent the SNP effect changes with age, instrument or with the rater of the behavior. If, on the other hand, no robust associations are uncovered, it will allow us to put a much tighter upper bound on the expected effect sizes for common variants, and tight bounds on the variance in effect size found over age and or rater. Therefore, the results will have scientific merit regardless of the outcome.

Data analysis methods:

Genome-wide Association analysis of each phenotype will take place at Duke University. The model for association testing is as follows:

$$Phenotype = \beta_0 + \beta_1 * SNP + \beta_2 * Z(age) + \beta_3 * sex + \beta_{4-9} * PC's + e$$

Covariates coding

- sex (coded 0 = F, 1 = M)
- z-score of age at the time of assessment
- The first 5 principal components.

Association is tested on the 22 autosomes only. For E-Risk, we will implement GCTA (http://cnsgenomics.com/software/gcta/mlmassoc.html) to conduct GWAs with related individuals (if the sample contains a large number of related subjects, such as twins or siblings. If this is not the case, a far simpler procedure can be used). A mixed effects procedure that includes 2 GRM's will be used. Briefly, the procedure is as follows:

1. Compute a Genetic relatedness matrix (GRM) based on genotyped SNPs (if this is difficult due to multiple genotype platforms being used in your cohort see: Fedko et al. (2015) for an alternative). In computing this GRM omit SNPs that fail HWE (p< 1e-6), have low MAF (<1%) or a low call rate (<99%).
2. Change the GRM into a GRM containing the pairwise genetic relationships above 0.05 (changing the values below 0.05 to 0). This GRM will form the first GRM in the mixed model.
3. Compute a series of GRMs each of which omits a single chromosome. These will form the remaining GRMs for the mixed model.
4. Run a mixed model in which the SNP is associated with the phenotype, while the model accounts for 1) the covariates defined above, and 2) the random effects associated with the 2 GRMs.

Meta-analysis of results of all samples will be carried out by Hill Fung Ip and Koen Bolhuis and be supervised by Michel Nivard at Free University, Amsterdam. The meta-analysis will take the form of a multivariate meta regression. We will apply genomic control based on the LDscore intercept (Bullik-Sullivan et al. 2014) and the appropriate marker filters at this stage.

Variables needed at which ages:

Study: E-Risk

Age 5

AGGET5: Aggressive Behaviour, Teacher Report form

AGGEM5: Aggressive Behaviour, Child Behavior Checklist parent

Age 7

TAGEE7: Age at Interview

AGGET7: Aggressive Behaviour, Teacher Report form

AGGEM7: Aggressive Behaviour, Child Behavior Checklist parent

Age 10

TAGEE10: Age at Interview

AGGET10: Aggressive Behaviour, Teacher Report form

AGGEM10: Aggressive Behaviour, Child Behavior Checklist parent

Age 12

TAGEE12: Age at interview

AGGET12: Aggressive Behaviour, Teacher Report form

AGGEM12: Aggressive Behaviour, Child Behavior Checklist parent
Age 18
TAGEE18: Age at Interview
CDSXE18: Conduct problem symptoms, self report
SR_INSUM18E: Attention problems, self report DSM-IV/V
Attention problems, Informant report DSM-IV/V

Study: Dunedin

TANTIS5	Teacher Rutter antisocial @ 5
PANTIS5	Parent Rutter antisocial @5
TANTIS7	Teacher Rutter antisocial @ 7
PANTIS7	Parent Rutter antisocial @7
TANTIS9	Teacher Rutter antisocial @ 9
PANTIS9	Parent Rutter antisocial @9
TANTIS11	Teacher Rutter antisocial @ 11
PANTIS11	Parent Rutter antisocial @11
TANTI511	mean Teacher Rutter antisocial, 5 to 11
PANTI511	mean parent Rutter antisocial, 5 to 11
FGHT511T	fighting - mean teacher rep 5 to 11
FGHT511P	mean parent report of fighting, 5 to 11
PQSA13	PQ socialized aggression @ 13, parent report
PQSA15	PQ socialized aggression @ 15, parent report
MPQAGR18	MPQ Aggression @ 18
INATTN9	inattn scale at 9, parents + teachers
INATTN11	inattn scale at 11, parents + teachers, RutterBeh
PQAT13	PQ inattention @ 13, parent report
PQAT15	PQ inattention @ 15, parent report

For both studies:
Genome-Wide SNP genotyping data, imputed 1000Gv3.
Sex
PCs 1-5 from PCA of genetic data

Significance of the Study (for theory, research methods or clinical practice):

This study will lead to an increased understanding of the genetics of externalizing behavior problems in children and adolescents, with a focus on aggressive and attention problems.

The proposed multivariate meta-analysis has statistical-theoretical and clinical significance. This study will be the first genome wide meta-analysis which explicitly includes both repeated measures at multiple ages and multiple behavioral raters.

Genome wide association studies almost exclusively consider univariate phenotypes. This has meant that a large number of observations which are available (multiple raters at multiple ages) have been omitted when performing GWAS. Given that twin studies find evidence for substantial

shared genetic effects between raters and across ages, discarding these repeated measures form analyses results in a loss in power. The inclusion of these repeated measures and measures obtained from multiple raters can clearly increase power for genetic locus (SNP) detection.

The theoretical significance lies in the handling of data from multiple informants. Twin studies have found evidence for rater specific (e.g. father, mother teacher or self-ratings) genetic variation in childhood aggression and attention problems (Areseneault, 2000) as well as age specific genetic variation (Kan et al 2014). These nuances have yet to be considered in the context of genome wide association studies (Middeldorp et al. 2016; Pappa et al., 2016). The benefit of considering rater differences in the context of a GWAS is that the results can be used to detect SNPs that differ across raters, and that we can take these SNPs forward in e.g. studies of adult psychopathology. Do SNPs unique to father ratings, mother ratings or teacher ratings of aggressive behavior differ in their correlations with adult psychopathology? Are specific raters more informative of the genetic liability of entirely different adult outcomes, or are the differences a matter of degree? Likewise, we will detect SNPs that influence the ratings of 2 or 3 raters. The same rationale exists for assessing the SNP x age interaction. Results from these part of the analyses will allow us to estimate the genetic correlation between an adult outcome (say bipolar disorder) and aggressive behavior measured at age 3, 7, 12 ,14,and 16.

The findings are informative for clinical practice where different raters are relied on in diagnosis and prognosis, and to inform on the need for intervention. Because of requirements for large sample sizes, up till now the first GWAS papers of childhood problems have had to pool data from different ages/raters and we hope we can now improve in the next set of analyses.

References cited:

Arseneault, L., Moffitt**,** T.E., Caspi, A., Taylor, A., Rijsdijk, F., Jaffee, S, Ablow, JC, & Measelle, J.R. (2003). Strong genetic effects on cross-situational antisocial behavior among 5-year-old children, according to mothers, teachers, examiners, and twin's self-reports. *Journal of Child Psychiatry and Psychology, 44, 832-848.*

Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., ... & Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics, 47(3), 291-295.

Fedko, I. O., Hottenga, J. J., Medina-Gomez, C., Pappa, I., van Beijsterveldt, C. E., Ehli, E. A., ... & Middeldorp, C. M. (2015). Estimation of Genetic Relationships Between Individuals Across Cohorts and Platforms: Application to Childhood Height. *Behavior genetics*, *45*(5), 514-528

Kan, K. J., Dolan, C. V., Nivard, M. G., Middeldorp, C. M., van Beijsterveldt, C. E., Willemsen, G., & Boomsma, D. I. (2013). Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(1), 12-25.

Middeldorp CM, Hammerschlag AR, Ouwens KG, Groen-Blokhuis MM, St Pourcain B, Greven CU & Boomsma DI. A Genome-Wide Association Meta-Analysis of Attention-Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts. J Am Acad Child Adolesc Psychiatry. 2016, 55(10):896-905.

Nivard et al. Genetic Overlap Between Schizophrenia and Developmental Psychopathology: Longitudinal and Multivariate Polygenic Risk Prediction of Common Psychiatric Traits During Development. *Schizophr Bull.* 2017 epub. doi: 10.1093/schbul/sbx031

Pappa, I., St Pourcain, B., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M. G., ... & Evans, D. M. (2015). A genome-wide approach to children's aggressive behavior: The EAGLE consortium. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 2015

Data Security Agreement

A

Provisional Paper Title	GWAS Meta-Analysis of Aggressive behavior and Attention Problems
Proposing Author	Moffitt, Sugden, Caspi
Today's Date	12 April 2017

Please keep one copy for your records and return one to the PI Sponsor

Please initial your agreement

xx	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
x	My project is covered by Duke or Otago IRB OR I have /will obtain IRB approval from my home institution.
x	I will treat all data as "restricted" and store in a secure fashion.
x	I will not share the data with anyone, including students or other collaborators not specifically listed on this concept paper.
x	I will not post data online or submit the data file to a journal for them to post. *Some journals are now requesting the data file as part of the manuscript submission process. The Dunedin Study cannot be shared because the Study Members have not given informed consent for unrestricted open access. Speak to Terrie or Avshalom for strategies for dealing with data sharing requests from Journals.*
x	Before submitting my paper to a journal, I will submit my draft manuscript and scripts for data checking, and my draft manuscript for co-author mock review, allowing three weeks
x	I will submit analysis scripts and new variable documentation to project data manager after manuscript gets accepted for publication.
x	I will return all data files to the Data Manager after the project is complete. Collaborators and graduates of DPPP may not take a data file away from the DPPP office. The data remains the property of the Study and cannot be used for further analyses without express, written permission.

Signature:

_____**Temi, Av, and Karen**_____

CONCEPT PAPER RESPONSE FORM

A

Provisional Paper Title	GWAS Meta-Analysis of Aggressive behavior and Attention Problems
Proposing Author	Moffitt, Sugden, Caspi from Duke
Other Contributors	Nivard, Bartels, Boomsma from Free University Amsterdam Richie Poulton from Otago Univ., Dunedin Study director. Louise Arseneault from King's College, E-Risk director.
Potential Journals	To be decided by the Amsterdam meta-analysis team
Today's Date	20 April 2017
Intended Submission Date	During 2017

Please keep one copy for your records and return one to the proposing author

B. To be completed by potential co-authors:

	Approved
	Not Approved
	Let's discuss, I have concerns

Comments:

Please check your contribution(s) for authorship:

	Conceptualizing and designing the longitudinal study
	Conceptualizing and collecting one or more variables
	Data collection
	Conceptualizing and designing this specific paper project
	Statistical analyses
	Writing
	Reviewing manuscript drafts
	Final approval before submission for publication
	Acknowledgment only, I will not be a co-author

Signature: _____