

# General Psychopathology in Children

Epidemiological Studies of Biological Mechanisms



# **GENERAL PSYCHOPATHOLOGY IN CHILDREN**

Epidemiological Studies of Biological Mechanisms

Alexander Neumann

ISBN: 978-94-6323-687-4

© Alexander Neumann, 2019

Chapter III.A,B,C;V.B: Copyright of these chapters has been transferred to the respective journals (see page 5). All rights reserved.

Chapter V.A is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>)

All other chapters are licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<https://creativecommons.org/licenses/by-sa/4.0/>).

Cover:

Yi Lin Shih (design) and Yu-Chin Her (painting)

DTI image is a derivative of "Webs'r'us" by jgmarcelino (<https://www.flickr.com/people/28755382@N00>) used under Creative Commons Attribution 2.0 Generic (<https://creativecommons.org/licenses/by/2.0/deed.en>)

Layout:

Daniel Iglesias González

Printing:

Gildeprint, Enschede, the Netherlands

# **General Psychopathology in Children**

Epidemiological Studies of Biological Mechanisms

Thesis

to obtain the degree of Doctor from the  
Erasmus University Rotterdam  
by command of the  
Rector Magnificus  
Prof. Dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board  
The public defense shall be held on Wednesday 21 June  
2019 at 9:30

by  
**Alexander Neumann**  
born in Berlin, Germany

**Erasmus University Rotterdam**



## **Doctoral Committee**

<b>Promotors</b>	Prof. Dr. H. Tiemeier Prof. Dr. M. van IJzendoorn Prof. Dr. M. Bakermans-Kranenburg
<b>Other members</b>	Dr. F. Rivadeneira Prof. Dr. B. Lahey Prof. Dr. J. Ormel

<b>Paranimfen</b>	Charlotte Cecil Andrea Cortes Hidalgo
-------------------	--

## MANUSCRIPTS THAT FORM THE BASIS OF THIS THESIS

Zondervan-Zwijnenburg, M.A.J., Veldkamp, S.A.M. , Nelemans, S.A., Neumann, A., Barzeva, S., Branje, S. J. T., ... & Boomsma, D.I. (2019). Parental Age and Offspring Childhood Mental Health: A Multi-Cohort, Population-Based Investigation. *Child Development* (in press). (Chapter III.A)

Sallis, H., Szekely, E., Neumann, A., Jolicoeur-Martineau, A., van IJzendoorn, M., ... & Evans J. (2019). General psychopathology, internalising and externalising in children and functional outcomes in late adolescence. *Journal of Child Psychology and Psychiatry* (in press). (Chapter III.B)

Neumann, A., Pappa, I., Lahey, B. B., Verhulst, F. C., Medina-Gomez, C., Jaddoe, V. W., ... & Tiemeier, H. (2016). Single nucleotide polymorphism heritability of a general psychopathology factor in children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(12), 1038-1045. (Chapter III.C)

Neumann, A., Nolte, I., Pappa, I., Pettersson, E., Rodriguez, A., ... & Tiemeier, H. (2019). A genome-wide association study of total child psychiatric problems scores (in preperation). (Chapter III.D)

Neumann, A., Muetzel, R.L., Lahey, B.B., Bakermans-Kranenburg, M.J., van IJzendoorn, M.H., ... & Tiemeier, H. (2019). White matter microstructure and the general psychopathology factor in children (submitted). (Chapter III.E)

Neumann, A., Walton, A., Alemany, S., Cecil, C.; Barker, E., ... & Tiemeier, H. (2019). ADHD symptoms and DNA methylation at birth and school-age (in preperation). (Chapter IV).

Neumann, A., Noppe, G., Liu, F., Kayser, M., Verhulst, F. C., Jaddoe, V. W., ... & Tiemeier, H. (2017). Predicting hair cortisol levels with hair pigmentation genes: a possible hair pigmentation bias. *Scientific reports*, 7(1), 8529. (Chapter V.A)

Neumann, A., Direk, N., Crawford, A. A., Mirza, S., Adams, H., Bolton, J., ... & Milaneschi, Y. (2017). The low single nucleotide polymorphism heritability of plasma and saliva cortisol levels. *Psychoneuroendocrinology*, 85, 88-95. (Chapter V.B)

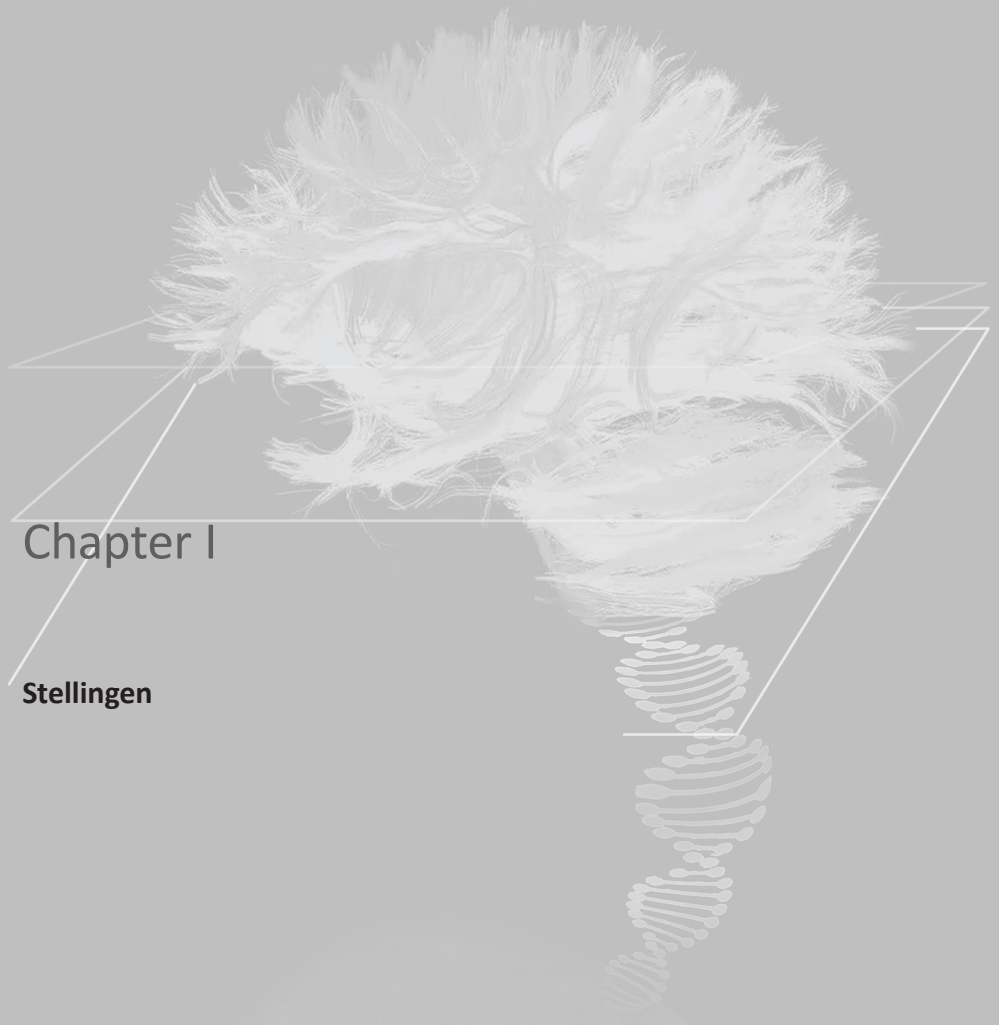


## TABLE OF CONTENTS

I.	Stellingen	9
II.	General Introduction	13
III.	Predictors, causes and outcomes of general and specific psychopathology	23
	A. Parental age and offspring childhood mental health: a multi-cohort, population-based Investigation	25
	B. The general psychopathology factor: An examination of the structure of child psychopathology across multiple cohorts	65
	C. Single nucleotide polymorphism heritability of a general psychopathology factor in children	95
	D. A genome-wide association study of total child psychiatric problems scores	119
	E. White matter microstructure and the general psychopathology factor in children	157
IV.	ADHD symptoms and DNA methylation at birth and school-age	211
V	Cortisol Genetics	237
	A. Predicting hair cortisol levels with hair pigmentation genes: a possible hair	239
	B. The low single nucleotide polymorphism heritability of plasma and saliva cortisol levels	275
VI.	General Discussion	297
VII.	Summary/Samenvatting	315
VIII	Appendix	323







Chapter I

Stellingen





## STELLINGEN

1. The co-occurrence of child psychiatric symptoms can be well explained by an underlying general psychopathology factor derived from multiple informants.
2. The general psychopathology factor is genetically heritable and associated with variation in common single nucleotide polymorphisms.
3. General psychopathology is related to lower global levels of white matter integrity, whereas specific externalizing levels are related to higher integrity.
4. DNA methylation at birth is associated with the development of ADHD symptoms.
5. Ethnicity-related stress cannot be studied with hair cortisol, as concentrations are related to hair color and structure.
6. Psychiatric epidemiological research data is almost never missing completely at random, therefore complete case analysis should be avoided.
7. Introspection of our conscious experience is not infallible (Dennett, 1988).
8. Psychiatric symptoms may be an adaptation, but this does not make them less problematic.
9. Failure to replicate is often blamed on study heterogeneity, yet the lack of power in the discovery is typically the main culprit.
10. Modern psychiatric epidemiology is the study of small effect sizes.
11. Free and open source software promotes collaboration, reproducibility and transparency. It therefore should be chosen over propriety software in science whenever possible.





Chapter II

Introduction

## INTRODUCTION

Psychiatric research in the last decades greatly illuminated the role of genetics, epigenetics, hormones and brain processes in psychiatric disorders. At the same time a wealth of research on the phenotypic level has shown that co-occurrence of psychiatric symptoms from different domains is pervasive. For example, behavioral and emotional problems correlate with a correlation coefficient of around 0.5 and half of patients with a psychiatric diagnosis have a second diagnosis.<sup>1,2</sup> However, the biology of the co-occurrence is less well understood and will be the theme of this thesis.

Perhaps the lack of study of co-occurrence in biological psychiatry is the result of biases, distorting our understanding of biology and impact the way we conduct research, arguably more than that of environmental processes. In the case of genetics Dar-Nimrod and Heine<sup>3</sup> discussed the following biases (adapted here to a psychiatric perspective): 1. psychiatric traits are the results of single genes, 2. genes deterministically impact the occurrence of a psychiatric disorder i.e. carriers of risk variants are guaranteed to have the disorder, 3. if a disorder is genetic, there are no other causes 4. heritability of a psychiatric disorder implies, that those at genetic risk form a homogeneous and distinct group. 5. heritability of a trait implies that it is naturally occurring and not an artificial construct.

Many of these biases are being addressed successfully in current psychiatric genetic research. For example, psychiatric genetics is not dismissing the role of other causes, as twin research shows that all psychiatric disorders have some proportion of non-genetic causes, for many disorders constituting the majority of effects.<sup>4</sup> Furthermore, the increasing use of polygenic scores, that predict levels of psychopathology based on hundreds to millions of SNPs, is reflecting the observation that psychiatric disorders are complex genetic disorders, which are influenced by many genetic variants.<sup>5</sup> Researchers also acknowledge that the environment can reduce the risk of developing a disorder either by compensating the genetic risk, or by interacting with risk effects as proposed by a diathesis-stress or differential susceptibility models: the degree to which a genetic variant affects a person is dependent on the presence of environmental circumstances.<sup>6</sup>

However, psychiatric genetics is still biased towards classification of distinct homogeneous groups. Most GWASs follow a case-control design in which the question is: does the frequency of a genetic variant change the odds of having a disorder or not, thus implying that a genetic variant would contribute to separation of people into two distinct groups, for example, those with and without ADHD<sup>7</sup>. While oversampling participants with diagnoses may make analyses more powerful by increasing contrasts, the lack of accounting for degree of symptom number or severity fails to capture the nature of psychopathology<sup>8</sup> and has undesired statistical consequences<sup>9</sup>. While there is an increase in GWAS studies of dimensional assessments, e.g. also of ADHD<sup>10</sup>, thus acknowledging that genetic risk may gradually increase or decrease the number and in-

tensity of symptoms, another classification bias is still at play. GWAS of single disorders or single domains assume that genetic variants increase the risk of a specific disorder/domain only. However, the possibility also exists that many genetic variants increase the risk of developing any psychiatric symptom, i.e. these variants would increase levels of general psychopathology. General psychopathology here, however, does not imply that people with the same levels of general psychopathology will necessarily have the same set of symptoms. Thus, carriers of genetic risk for general psychopathology may not form a homogeneous group with the same symptoms and thus do not follow the implicit expectation of a genetic disorder. In this scenario, research method would require adjustment to measure and jointly analyze a broad set of symptoms.

This bias of attempting to find etiological factors which cause distinct diagnoses or narrow sets of symptoms instead of general psychopathology is not exclusive to genetic studies. Most biological studies focus on the analysis of single disorders or psychopathology domains at a time, whether it be neuroimaging or psychoendocrinological studies, despite evidence that neural and endocrine features are associated with multiple psychopathology domains and psychological variables in general. For instance, global white matter integrity is associated with cognitive abilities<sup>11</sup>, depression<sup>12</sup>, attention and internalizing problems<sup>13</sup>; cortisol levels were associated with post-traumatic stress disorder, schizophrenia, bipolar disorder<sup>14,15</sup> and treatment response to depression<sup>16</sup>. Yet, systematic investigations of general psychopathology are lacking in biological psychiatry.

The main question of this dissertation is: which biological factors are associated with child psychopathology in general and which biological factors are specific to certain psychopathology domains? Before discussing how to separate general from specific effects, it is necessary to first introduce the psychopathology domains will be studied in this thesis. The most commonly studied domains in children are the internalizing, externalizing and attention disorders. Internalizing disorders include anxiety and depressive symptoms, whereas externalizing disorders consist of aggression and rule-breaking behaviors. Attention problems, especially at young age, are sometimes defined as externalizing, but there is evidence that they should be regarded as a separate domain in later school age.<sup>17</sup>

General psychopathology can be investigated in several ways. One is the use of traditionally defined domain scores, such as internalizing and externalizing scores, followed by comparisons whether effects on these psychopathology scores are similar between the domains. However, if truly general effects are at play, then the associations with single domains may be downward biased compared to measures of general psychopathology, as each domain score would be an incomplete measure of general psychopathology. The simplest alternative is the use of a total sum of psychiatric symptoms scores. The advantage of this approach is the easy computation and interpretability of the score. However, it may not be the best representation of general psychopathology, as it assumes that all symptoms are equally affected by general psychopathology and it



does not take into account correlation between the symptoms nor between the general and specific factors.<sup>18</sup> A more sophisticated approach has therefore been the use of latent variables models to specify both general and specific psychopathology factors simultaneously.<sup>19–21</sup> In these bifactor models symptoms are hypothesized to be caused by a general psychopathology factor, as well as domain specific factors. These models can be extended to include multiple informants, reducing the chance that rater bias would inflate levels of general psychopathology. Relating the factors derived from a bifactor model to predictors or outcomes allows the testing of general and specific effects on/ of psychopathology.

All three approaches will be used in this thesis, with the individuals studies described in Chapter III-V. Chapter III attempts to differentiate which (mostly) biological factors associate with psychopathology in general and which factors with specific domains. Chapter IV focuses on one particular disorder: attention-deficit and hyperactivity disorder. Chapter V concludes with investigations into the stress hormone cortisol, which is believed to be causally involved in the development of psychiatric symptoms.

Chapter III consists of five studies investigating various potential predictors, causes and outcomes of general and specific psychopathology. The first study “Parental age and offspring childhood mental health: a multi-cohort, population-based Investigation” focuses on the beginning of life and discusses the age of parents at delivery and the risk of the child to develop psychiatric symptoms. It is well established that higher maternal age is associated with heightened risk of pregnancy complications and health problems in the offspring, with some evidence for also adverse effects of higher paternal age.<sup>22–24</sup> This raises the question, whether the same is true for mental health, and if so, whether the effects are stronger for internalizing or externalizing problems, or the same.

As mentioned above, using only scores of individual domains may not be the best approach for disentangling general and specific effects. In the second study “The general psychopathology factor: An examination of the structure of child psychopathology across multiple cohorts” we therefore introduce a bifactor model of general and specific psychopathology. In this study we attempt to find a common structure of psychopathology in school-aged children among three different cohorts. Furthermore, we compare unifactor and bifactor structures in their ability to predict adult performance and mental health outcomes.

In the next paper “Single nucleotide polymorphism heritability of a general psychopathology factor in children”, we continue using latent factor models to determine the single nucleotide polymorphism (SNP) heritability of general psychopathology. SNP heritability refers to the variance explained by the additive effects of common genetic variants across the genome. Knowing the magnitude of the SNP heritability is interesting as individual SNPs typically have very small effect sizes. Thus the joint effect of all variants

across the genome, typically represented by a half million markers or more, is more informative of the overall heritability of a trait than the top associated SNPs.

While the total SNP heritability gives an important perspective on the overall contributions of SNPs, it is also important to detect the specific genetic loci associated with general psychopathology to improve understanding of etiology and for the detection of treatment targets. An approach to detect specific loci is to associate each SNP separately with an outcome in a genome-wide association study (GWAS). As a follow up to the SNP heritability study we therefore perform a GWAS of a total psychiatric sum score, as proxy for general psychopathology.

The last study in the first chapter revisits the bifactor models introduced in the previous studies, however, this time the general and specific psychopathology factors are related to white matter integrity. White matter is essential for efficient communication between brain regions and variations in microstructure may be associated with the presence and severity of psychiatric symptoms. Specifically, Zald et al.<sup>25</sup> hypothesized that global white matter microstructure differences across the whole brain are related to variability in general psychopathology, whereas variation in specific region causes specific symptoms. We test this hypothesis in school-aged children.

Chapter IV presents an epigenetic approach to further our biological psychiatric understanding. A growing number of research investigates variations in DNA methylation in relation to psychopathology. DNA methylation is influenced by genetic and environmental factors and has the potential to impact gene expression. It is therefore an interesting potential mediator of genetic and environmental risks or biomarker for adverse exposures. Similar to a GWAS it is possible to associate DNA methylation at hundreds of thousands of CpG sites with psychiatric symptoms. The first EWASs of psychiatric symptoms are being performed, however, large multi-center consortia efforts are lacking. We present a prospective meta-analytic EWAS on ADHD, a common childhood disorder. Unlike the genome, the epigenome varies over time and thus assessment time becomes important. We therefore associate DNA methylation both at birth and at school-age with ADHD symptoms and compare results.

The final chapter revolves around the stress hormone cortisol. Cortisol is a hormone, that is released in reaction to both physical and psychological stress.<sup>26–28</sup> Cortisol may also be involved in the etiology of psychopathology, as cortisol injections increase depressive behavior in animal models<sup>29</sup> and alterations in baseline levels are associated with some disorders in humans.<sup>14,15</sup> However, as cortisol is a highly dynamic hormone, not only responding to external stimuli, but also showing a diurnal rhythm<sup>30</sup>, and an excretion as pulse pattern<sup>31</sup>, finding the optimal cortisol assessment method has been challenging in psychiatric research.

The first study in this chapter investigates the utility of measuring cortisol in hair samples. Cortisol accumulates in hair and provides a more long-term profile of cortisol exposure. However, some research suggested that cortisol levels are related to hair color, though, it is difficult to distinguish to which degree this effect is due to hair col-

or or ethnicity/race. We attempted to disentangle the association by investigating the independent contributions of genetically determined hair color and genetic ancestry.

In the last study, we examine the genetics of acute cortisol levels in blood and saliva. Several studies investigated the heritability of cortisol using known family relationships to infer genetic effects,<sup>32–35</sup> however, molecular studies are lacking. We therefore estimate and compare the SNP heritability of various acute cortisol measures.

## RESEARCH SETTING

The primary focus of this dissertation is the identification of determinants and consequences of general psychopathology in children. As it would be unethical to randomize potential risk factors of psychopathology, we employ various epidemiological methods in large observational studies to study the causes of general and specific psychopathology. General psychopathology as defined here is a dimensional construct and the general population therefore displays varying degrees of it with no clear threshold for a disordered status. Therefore all the presented studies describe the general population and the whole range of general psychopathology.

The majority of the studies in this dissertation were conducted within consortia of many institutions and present the combined results of several cohorts. The study of parental age was embedded in the consortium of individual development (<https://individualdevelopment.nl/>) and included four Dutch cohorts. The study about the structure of psychopathology is the first DREAM BIG collaboration (<http://dreambigresearch.com/>) and comprises Canadian, British and Dutch cohorts. The GWAS on a total child psychiatric problem score is based on the results of 16 cohorts from North America, Europe and Australia from the EARly Genetics and Lifecourse Epidemiology (EAGLE) consortium. Finally, the CORNET<sup>36</sup> consortium consisting of cohorts from Europe and the US contributed substantially to most analyses in the SNP heritability of cortisol paper.

Except for the latter, all studies involved the Generation R cohort. Generation R is a population-based birth cohort based in Rotterdam, the Netherlands.<sup>37</sup> Expecting mothers with a delivery date from 2002 to 2006 were invited to participate in this study. The parents and later their children's characteristics and development were assessed from birth. At the time of writing, the most recent assessment wave is at the age of 13 years. However, this thesis largely focuses on the early school-ages (6 to 10 years). This is an interesting period to study general psychopathology. Several disorders do not reach substantial incidence levels until puberty, but varying levels of general psychopathology may be already present and manifest in various disorders in childhood and later life. The study of general psychopathology in childhood is therefore likely of high relevance and I hope that the following chapters will contribute to our understanding of the etiology and biological correlates of general psychopathology.

## REFERENCES

1. Achenbach TM, Ivanova MY, Rescorla LA, Turner L V., Althoff RR. Internalizing/Externalizing Problems: Review and Recommendations for Clinical and Research Applications. *J Am Acad Child Adolesc Psychiatry*. 2016;55(8):647-656. doi:10.1016/j.jaac.2016.05.012
2. Newman DL, Moffitt TE, Caspi A, Silva PA. Comorbid mental disorders: Implications for treatment and sample selection. *J Abnorm Psychol*. 1998;107(2):305-311.
3. Dar-nimrod I, Heine SJ. Genetic Essentialism : On the Deceptive Determinism of Genetic. *Psychol Bull*. 2010;137(5):800-818. doi:10.1037/a0021860
4. Polderman TJC, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015;47(7):702-709. doi:10.1038/ng.3285
5. Jansen PR, Polderman TJC, Bolhuis K, et al. Polygenic scores for schizophrenia and educational attainment are associated with behavioural problems in early childhood in the general population. *J Child Psychol Psychiatry*. 2018;59(1):39-47. doi:10.1111/jcpp.12759
6. Belsky J, Bakermans-Kranenburg MJ, van IJzendoorn MH. For Better and For Worse. *Curr Dir Psychol Sci*. 2007;16(6):300-304. doi:10.1111/j.1467-8721.2007.00525.x
7. Demontis D, Walters RK, Martin J, et al. Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv*. 2017:145581. doi:10.1101/145581
8. Krueger RF, Kotov R, Watson D, et al. Progress in achieving quantitative classification of psychopathology. *World Psychiatry*. 2018;17(3):282-293. doi:10.1002/wps.20566
9. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods*. 2002;7(1):19-40. doi:10.1037/1082-989X.7.1.19
10. Middeldorp CM, Hammerschlag AR, Ouwens KG, et al. 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. doi:10.1016/j.jaac.2016.05.025
11. Muetzel RL, Mous SE, van der Ende J, et al. White matter integrity and cognitive performance in school-age children: A population-based neuroimaging study. *Neuroimage*. 2015;119:119-128. doi:10.1016/j.neuroimage.2015.06.014
12. Shen X, Reus LM, Cox SR, et al. Subcortical volume and white matter integrity abnormalities in major depressive disorder: findings from UK Biobank imaging data. *Sci Rep*. 2017;7(1):5547. doi:10.1038/s41598-017-05507-6
13. Loe IM, Lee ES, Feldman HM. Attention and Internalizing Behaviors in Relation to White Matter in Children Born Preterm. *J Dev Behav Pediatr*. 2013;34(3):156-164. doi:10.1097/DBP.0b013e3182842122
14. Girshkin L, Matheson SL, Shepherd AM, Green MJ. Morning cortisol levels in schizophrenia and bipolar disorder: A meta-analysis. *Psychoneuroendocrinology*. 2014;49(1):187-206. doi:10.1016/j.psyneuen.2014.07.013
15. Yehuda R, Seckl J. Minireview: Stress-Related Psychiatric Disorders with Low Cortisol Levels: A Metabolic Hypothesis. *Endocrinology*. 2011;152(12):4496-4503. doi:10.1210/en.2011-1218
16. Fischer S, Strawbridge R, Herane Vives A, Cleare AJ. Cortisol as a predictor of psychological therapy response in depressive disorders: systematic review and meta-analysis. *Br J Psychiatry*. 2016. doi:10.1192/bjp.bp.115.180653

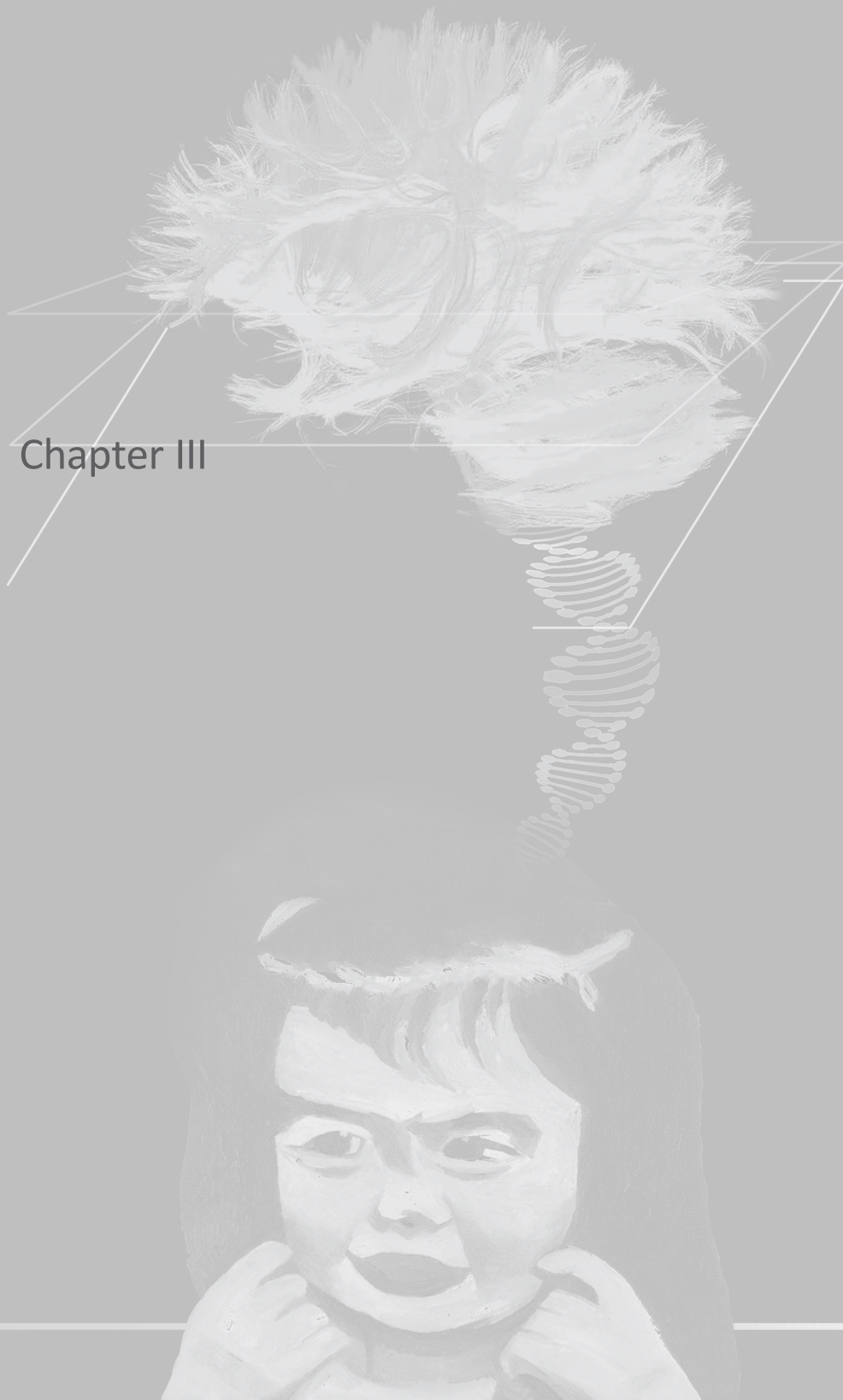
17. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms and Profiles. . (University of Vermont Youth, and Families RC for C, ed.). Burlington, VT; 2001.
18. Distefano C, Míndril D. Understanding and Using Factor Scores : Considerations for the Applied Researcher. *Pract Assessment, Res Eval.* 2009;14(20).
19. Caspi A, Houts RM, Belsky DW, et al. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci.* 2014;2(2):119-137. doi:10.1177/2167702613497473
20. Laceulle OM, Vollebergh W a. M, Ormel J. The Structure of Psychopathology in Adolescence: Replication of a General Psychopathology Factor in the TRAILS Study. *Clin Psychol Sci.* 2015:1-11. doi:10.1177/2167702614560750
21. Lahey BB, Krueger RF, Rathouz PJ, Waldman ID, Zald DH. A hierarchical causal taxonomy of psychopathology across the life span. *Psychol Bull.* 2017;143(2):142-186. doi:10.1037/bul0000069
22. Carslake D, Tynelius P, Berg G Van Den, Smith GD, Rasmussen F. Associations of parental age with health and social factors in adult offspring. *Methodological pitfalls and possibilities. Sci Rep.* 2017;7(45278):1-16. doi:10.1038/srep45278
23. Khandwala YS, Baker VL, Shaw GM, Stevenson DK, Lu Y, Eisenberg ML. Association of paternal age with perinatal outcomes between 2007 and 2016 in the United States: population based cohort study. *Bmj.* 2016;363(363). doi:10.1136/bmj.k4372
24. Hegelund ER. Interpregnancy Interval and Risk of Adverse Pregnancy Outcomes : A Register-Based Study of 328,577 Pregnancies in Denmark. *Matern Child Health J.* 2018;22(7):1008-1015. doi:10.1007/s10995-018-2480-7
25. Zald DH, Lahey BB. Implications of the Hierarchical Structure of Psychopathology for Psychiatric Neuroimaging. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2017;2(4):310-317. doi:10.1016/j.bpsc.2017.02.003
26. West DWD, Phillips SM. Associations of exercise-induced hormone profiles and gains in strength and hypertrophy in a large cohort after weight training. *Eur J Appl Physiol.* 2012;112(7):2693-2702. doi:10.1007/s00421-011-2246-z
27. Barton R, Stoner H, Watson S. Relationships among plasma cortisol, adrenocorticotrophin, and severity of injury in recently injured patients. *J Trauma.* 1987;27(4):384-392. [http://journals.lww.com/jtrauma/Abstract/1987/04000/Relationships\\_among\\_Plasma\\_Cortisol,.7.aspx](http://journals.lww.com/jtrauma/Abstract/1987/04000/Relationships_among_Plasma_Cortisol,.7.aspx). Accessed January 29, 2016.
28. Kudielka BM, Hellhammer DH, Wüst S. Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. *Psychoneuroendocrinology.* 2009;34(1):2-18. doi:10.1016/j.psyneuen.2008.10.004
29. Sterner EY, Kalynchuk LE. Progress in Neuro-Psychopharmacology & Biological Psychiatry Behavioral and neurobiological consequences of prolonged glucocorticoid exposure in rats: Relevance to depression. *Prog Neuropsychopharmacol Biol Psychiatry.* 2010;34(5):777-790. doi:10.1016/j.pnpbp.2010.03.005
30. Adam EK. Transactions among adolescent trait and state emotion and diurnal and momentary cortisol activity in naturalistic settings. *Psychoneuroendocrinology.* 2006;31(5):664-679. doi:10.1016/j.psyneuen.2006.01.010
31. Young E, Abelson J, Lightman S. Cortisol pulsatility and its role in stress regulation and health. *Front Neuroendocrinol.* 2004;25(2):69-76. doi:10.1016/j.yfrne.2004.07.001
32. Froehlich JC, Zink RW, Li TK, Christian JC. Analysis of heritability of hormonal responses to al-

- cohol in twins: beta-endorphin as a potential biomarker of genetic risk for alcoholism. *Alcohol Clin Exp Res*. 2000;24(3):265-277. <http://www.ncbi.nlm.nih.gov/pubmed/10776662>.
33. Inglis GC, Ingram MC, Holloway CD, et al. Familial pattern of corticosteroids and their metabolism in adult human subjects - The Scottish adult twin study. *J Clin Endocrinol Metab*. 1999;84(11):4132-4137.
  34. Meikle a W, Stringham JD, Woodward MG, Bishop DT. Heritability of variation of plasma cortisol levels. *Metabolism*. 1988;37(6):514-517. <http://www.ncbi.nlm.nih.gov/pubmed/2967419>.
  35. Wüst S, Federenko I, Hellhammer DH, Kirschbaum C. Genetic factors, perceived chronic stress, and the free cortisol response to awakening. *Psychoneuroendocrinology*. 2000;25(7):707-720. doi:10.1016/S0306-4530(00)00021-4
  36. Bolton JL, Hayward C, Direk N, et al. Genome wide association identifies common variants at the SERPINA6/SERPINA1 locus influencing plasma cortisol and corticosteroid binding globulin. *PLoS Genet*. 2014;10(7):e1004474. doi:10.1371/journal.pgen.1004474
  37. Kooijman MN, Kruithof CJ, van Duijn CM, et al. The Generation R Study: design and cohort update 2017. *Eur J Epidemiol*. 2016;31(12):1243-1264. doi:10.1007/s10654-016-0224-9





Chapter III









## Chapter III.A

### **Parental Age and Offspring**

### **Childhood Mental Health: A Multi-Cohort, Population-Based Investigation**

Zondervan-Zwijenburg, M.A.J.\* , Veldkamp,  
S.A.M. \*, Nelemans, S.A., Neumann, A.,  
Barzeva, S., Branje, S. J. T., van Beijsterveldt  
C.E.M., Meeus, W.H.J., Tiemeier, H., Hoijtink,  
H.J.A., Oldehinkel, A.J., Boomsma, D.I.

\*These authors contributed equally to this  
work

## ABSTRACT

To examine the contributions of maternal and paternal age on offspring externalizing and internalizing problems, this study analyzed problem behaviors at age 10-12 years from four Dutch population-based cohorts (N = 32,892) by a multiple informant design. Bayesian evidence synthesis was used to combine results across cohorts with 50% of the data analyzed for discovery and 50% for confirmation. There was evidence of a robust negative linear relation between parental age and externalizing problems as reported by parents. In teacher-reports, this relation was largely explained by parental socio-economic status. Parental age had limited to no association with internalizing problems. Thus, in this large population-based study, either a beneficial or no effect of advanced parenthood on child problem behavior was observed.



Since 1995, the mean maternal age at first birth has increased at a rate of 0.10 years per year in OECD countries, and in 2017 exceeded 30 years in the vast majority of these countries (Organisation for Economic Co-operation and Development, 2017). Only in Mexico was the mean age of women at childbirth lower than 28 years, and only in eight countries was it between 28 and 30 years of age. Women's reproductive years generally range from about 15 to 45 years.<sup>1</sup> Within this wide age range some periods are generally considered more suitable to have children than others, but which parental reproductive ages are optimal for offspring physical and mental health has been a matter of debate ever since individuals have engaged in active birth control. Whereas having children at an advanced age was quite common historically, when families tended to be larger<sup>2</sup>, the current trend to delay childbearing has given rise to public health concerns

### **Concerns Regarding Delayed Childbearing**

Concerns regarding delayed childbearing are understandable, as a large number of research reports highlight that increased maternal age at childbirth is associated with several adverse consequences, ranging from physical problems such as increased BMI, blood pressure and height<sup>3</sup> to psychiatric conditions such as autism<sup>4,5</sup>, bipolar disorder<sup>6</sup>, symptoms of depression, anxiety and stress<sup>7</sup>, and poor social functioning<sup>8</sup>. More recently, increased paternal age at birth has also been associated with adverse child outcomes, such as stillbirth and cleft palate.<sup>9</sup> In over 40 million live births between 2007 and 2016, having an older father increased the risk of low birthweight, apgar score, and premature birth.<sup>10</sup> A study of the Danish population, which included 2.8 million persons, found that older fathers are at risk of having offspring with intellectual disabilities, autism spectrum disorders and schizophrenia.<sup>11,12</sup>

Several, not mutually exclusive, mechanisms have been proposed to explain the increased physical and mental health risks in offspring of older parents. First, age-related deterioration of the functioning of women's reproductive organs, such as DNA damage in germ cells, and worse quality of oocytes and placenta, can increase the risk of obstetric and perinatal complications.<sup>13</sup> Second, male germline cells undergo cell replication cycles repeatedly during aging, with de novo point mutations accumulating over time<sup>14</sup> and the number of de novo mutations in the newborn increasing with higher age of the father at the time of conception<sup>15,16</sup>. Although weaker than with paternal age, de novo mutations in offspring correlate with maternal age as well.<sup>17,18</sup> Third, genomic regions in the male germline may become less methylated with increasing age and alter the expression of health-related genes.<sup>19</sup> Fourth, age effects can be due to selection, with older parents differing from younger ones in characteristics that are relevant for developmental outcomes in their offspring, such as poor social skills. The influence of selection effects can be exacerbated by assortative mating.<sup>20</sup> Fifth, being the child of older parents carries the risk of having to cope with parental frailty or losing a parent at a relatively young age,<sup>21</sup> and the stress evoked by these experiences may trigger health problems. Most of these mechanisms involve consequences of biological ageing. Par-

enthood at an advanced age is disadvantageous from a biological perspective; except for very young, physiologically immature mothers, younger parents are in a better physical condition.

### **Possible Benefits of Delayed Childbearing**

Whereas the effects of older parental age on children's physical health and psychiatric disorders tend to be predominantly negative, the effects of older parental age on mental health problems with a stronger psychosocial component, such as externalizing and internalizing problems, tend to be more inconsistent. An indication that the negative consequences of high parental age may stretch beyond clinical diagnosis is provided by Tearne and colleagues,<sup>7,22</sup> who found that high maternal age predicted symptoms of depression, anxiety and stress in daughters, and by Janecka and colleagues<sup>23</sup> who reported a negative association between advanced paternal age and social development. In contrast, in several population-based studies, offspring of older parents, particularly of older mothers, perform better at school and work, score higher on intelligence tests, report better health and higher well-being, use fewer drugs, and have fewer behavioral and emotional problems than offspring of younger parents.<sup>3,11,21,22,24,25</sup>

While the biology of ageing seems to put older parents in an unfavorable position with regard to their offspring's physical and mental health, these contradictory effects of parental age on offspring mental health outcomes might be explained by a psychosocial perspective. Being a child of older parents can have substantial benefits,<sup>26</sup> as older parents not only are often in a better socioeconomic position than young parents,<sup>27</sup> thereby providing a more favorable environment for children, they also have greater life experience. Furthermore, older parents display more hardiness<sup>28</sup> and tend to have less substance use and fewer mental health problems,<sup>29</sup> hence score higher on parenting factors that promote health and development.<sup>29,30</sup> In part, positive associations of advanced parental age could be related to selection effects. In young people, substance abuse and related externalizing problems go together with earlier sexual activity,<sup>31</sup> which increases the probability that intergenerational transmission of externalizing problems occurs at an early parental age.<sup>32</sup> Like age-related parental characteristics that may have negative effects on offspring outcomes, the influence of such selection effects can be exacerbated by assortative mating.<sup>20</sup>

In sum, whereas advanced parenthood, particularly advanced paternal age, has primarily been associated with physical health and neurodevelopmental outcomes, such as autism and schizophrenia, advanced parenthood, particularly advanced maternal age, rather seems to predict mental health problems with a stronger psychosocial component, such as externalizing problems. Although it seems plausible that parental age interferes with subclinical problems and traits underlying these conditions, comprehensive evidence from population-based cohorts is scarce and inconsistent, and more empirical evidence is desirable. Moreover, prior population-based studies that used continuous measures of mental health problems usually focused on cognitive or behavioral



problems<sup>3,25</sup> and, with a few exceptions that require replication in other cohorts<sup>7,22,23</sup> rarely included internalizing problems. A final reason to extent the research conducted thus far with the present study is the wide variety of populations, designs and outcomes used, which makes it hard to distinguish between substantive variation in association patterns and sample-specific artefacts. In short, there is a need for studies that investigate both maternal and paternal age effects on continuously assessed core dimensions of offspring mental health (including internalizing problems) and use robust analytical methods that allow the possibility of increased risk for both young and old parenthood.

### **The Present Study**

We investigated parental age effects on offspring externalizing and internalizing problems around age 10-13 years in four Dutch population-based cohorts: Generation R (Gen-R), the Netherlands Twin Register (NTR), the Research on Adolescent Development and Relationships-Young cohort (RADAR-Y), and the Tracking Adolescents' Individual Lives Survey (TRAILS) (see Table 1). The Netherlands is characterized by a high maternal age at birth, and relatively few teenage pregnancies. In 1950, 1.6% of the children were born to mothers younger than 20 years of age, with a comparable percentage (1.7%) in 1990. In 2016 this number had decreased to 0.6%. In contrast, the percentages of women who gave birth at an age above 40 years were 8.5% in 1950, 1.5% in 1990, and 4.3% in 2016.<sup>33</sup>

As the perception of childhood problems may differ for different informants,<sup>34,35</sup> we aimed to obtain a comprehensive set of outcome measures of internalizing and externalizing problems through a multiple informant design. The four cohorts provided reports from mothers, fathers, the children themselves, and the children's teachers. The addition of reports from teachers is particularly valuable, because their reports are unlikely to be affected by parental age-related report biases. We tested both linear and nonlinear effects, to be better able to distinguish effects of older parenthood versus younger parenthood. We tested effects with and without adjusting for child gender and socio-economic status. Socio-economic status was included as a covariate to get an impression of the relative importance of socio-economic factors in explaining parental age effects.

Bayesian evidence synthesis was used to summarize the results over the cohorts. The current era is one of increased awareness of the need for replication research before making scientific claims.<sup>36</sup> Therefore, in this study, the datasets of the four cohort studies were used to evaluate the same set of hypotheses with respect to the relation

between parental age and offspring mental health problems. This approach is called Bayesian evidence synthesis.<sup>37</sup>

## METHOD

### Participants

The participants in this study came from the Gen-R, NTR, RADAR-Y, and TRAILS population cohort studies. Table 2 gives the total sample size and information on parental age for each cohort. The total number of children in each cohort was 4,769 for Gen-R, 25,396 for NTR, 497 for RADAR-Y, and 2,230 for TRAILS.

Gen-R mothers were recruited in the city of Rotterdam during pregnancy. Their partners and later their children were also invited to participate. For Gen-R, participants from the child age-10 study wave (born between 2002 and 2006) were included if they had complete information on maternal age and a child behavioral problems sum score by at least one informant. When multiple children from one family were present, one sibling was randomly removed ( $N = 397$ ) to create a sample of unrelated individuals. Mean child age for mother report: 9.72 ( $SD = 0.32$ ), father report: 9.77 ( $SD = 0.32$ ), and child self-report: 9.83 ( $SD = 0.36$ ). 71.2% of the Gen-R sample is Dutch or European. Other groups are Suriname (6.4%), Turkish (5.3%), and Moroccan (4.2%). Mother's educational level is low (i.e., no education or primary education) for 9%, intermediate (i.e., secondary school, vocational training) for 42%, and high (i.e., bachelor's degree, university) for 49%. Based on CBCL T-scores for mother reports, 93.2% of the children had non-clinical scores for internalizing problems, 4.7% scored in the borderline category, and 2.1% scored in the clinical category. With respect to externalizing problems, 97.0% scored in the non-clinical category, 1.9% in the borderline category, and 1.0% in the clinical category.

The NTR study recruits new-born twins from all regions in the Netherlands. Here we included the data on 10-year-olds who were born between 1986 and 2007. Children were not included if they had a severe handicap which interfered with daily functioning. Mean child age for mother report was 9.95 ( $SD = 0.51$ ), father report 9.94 ( $SD = 0.50$ ) and teacher report 9.80 ( $SD = 0.58$ ). The children in NTR were mostly born in the Netherlands (99.5%). The remaining 0.5% consisted mainly of other West European nationalities (0.4%). Parents in the NTR were mostly born in the Netherlands (95.7% of fathers and 96.7% of mothers). 3.1% of mothers had a low skill occupation (primary education), 11.4% had an occupation that required lower secondary education, 40.3% had an upper secondary educational level, 30.6% had a higher vocational occupation level, and 14.6% worked at the highest (i.e. scientific) level. According to mother reports for internalizing problems, 86.1% of children had a non-clinical score, 5.9% had a borderline score,

and 8.0% scored in the clinical range. For externalizing problems, 85.7% scored in the non-clinical range, 6.5% scored in the borderline range, and 7.8% in the clinical range.

The RADAR-Y sample was recruited in the province of Utrecht and four large cities in the mid-west of the Netherlands. Because the RADAR-Y study had a focus on delinquency development, children with borderline externalizing behavior problems at age 12 were oversampled. All participants from the first wave of data collection, born between 1990 and 1995, were selected. The mean age of the children at this wave was 13.03 years (SD = 0.46). The sample consisted mainly of native Dutch (87.9%) children. Remaining participants belonged to the following groups: Surinam (2.4%), Indonesian/Moluccan (2.4%), Antillean (1.8%), Turkish (0.4%), and other (4.8%). The majority of children came from families with a medium or high socio-economic status (89.2%). According to the children's reports for externalizing problems, 81.6% of the participants had a non-clinical score, 7.2% had a borderline score, and 11.2% scored in the clinical range. Using the cutoff scores for the depression scale as described by Reynolds,<sup>38</sup> 4.0% of the children scored in the subclinical or clinical range of depressive symptoms. Using the cutoff scores for the anxiety scale of Birmaher et al.,<sup>39</sup> 5.3% of the children scored in the subclinical or clinical range for anxiety symptoms.

The TRAILS sample was recruited in the Northern regions of the Netherlands. All participants from the first wave of data collection (born between 1990 and 1991) were selected. The mean age of the children at the first wave was 11.09 (SD = 0.56). The large majority of participants were Dutch (86.5%), with other participants being Surinam (2.1%), Indonesian (1.7%), Antillean (1.7%), Moluccan (0.7%), Turkish (0.5%), and other (6.9%). Based on mother-reported sum-scores for the internalizing and externalizing scales, TRAILS participants were categorized in a non-clinical, borderline, or clinical category. For internalizing problems, 67.3% of the participants had a non-clinical score, 13.9% had a borderline score, and 18.8% had a clinical score. For externalizing problems, 74.5% had a non-clinical score, 10.2% a borderline score, and 15.4% had a score in the clinical range.

To summarize, the cohorts represented the entire Dutch geographic region across all strata from society. They had a similar distribution of SES. The percentage of participants with parents born in the Netherlands was relatively high in NTR (>95%), around 87% in Radar-Y and TRAILS and relatively low in Gen-R (<72%). The percentage of non-clinical behavioral problems was highest in Gen-R and lowest in TRAILS.

All studies were approved by central or institutional ethical review boards. The participants were treated in compliance with the Declaration of Helsinki, and data collection was carried out with their adequate understanding and parental consent. All measures in RADAR-Y were self-reports. In the other cohorts, children were rated by any combination of: their parents, themselves, or their teachers. Table 3 shows the total number of children in each cohort, and the number of participants with an externalizing



**Table 1:** General Cohort Information

Full cohort name	Short name	Website	Birthyears	References (DOI)
Generation R	Gen-R	generationr.nl	2002-2006	10.1007/s10654-016-0224-9 10.1016/j.jaac.2012.08.021
Netherlands Twin Register	NTR	tweelingenregister.org	1986-2017	10.1017/thg.2012.118 10.1016/j.jaac.2012.10.009
Research on Adolescent Development And Relationships – Young Cohort	RADAR-Y	www.uu.nl/onderzoek/radar	1990-1995	10.1111/cdev.12547 10.17026/dans-zrb-v5wp
TRacking Adolescents' Individual Lives Survey	TRAILS	trails.nl	1989-1991	10.1093/ije/dyu225

**Table 2:** Cohort Descriptive Statistics of Total Sample Size and Parental Age in Current Study

Cohort	N	Maternal age at birth child		Paternal age at birth child	
		Range	<i>M (SD)</i>	Range	<i>M (SD)</i>
Gen-R	4,769	16.56 – 46.85	31.68 (4.79)	17.61 – 68.67	34.24 (5.58)
NTR	25,396	17.36 – 47.09	31.35 (3.95)	18.75 – 63.61	33.76 (4.71)
RADAR-Y	497	17.80 – 48.61	31.38 (4.43)	20.34 – 52.52	33.70 (5.10)
TRAILS	2,23	16.34 – 44.88	29.32 (4.58)	18.28 – 52.09	31.99 (4.71)

**Table 3:** Total Sample Size and Sample Sizes per Informant per Cohort

	Gen-R (N= 4,769)		NTR (N=25,396)		RADAR-Y (N=497)		TRAILS (N=2,230)	
<b>Variable Informant</b>								
<b>Externalizing behavior problems</b>	Child	BPM <sup>a</sup> 4,01	-	-	YSR <sup>b</sup> 491	YSR <sup>b</sup> 2,188		
	Mother	CBCL <sup>c</sup> 4,549	CBCL <sup>c</sup> 21,921	-	-	CBCL <sup>c</sup> 1,965		
	Father	CBCL <sup>c</sup> 3,259	CBCL <sup>c</sup> 14,715	-	-	-		
	Teacher	-	TRF <sup>d</sup> 12,573	-	-	TCP <sup>e</sup> 1,925		
<b>Internalizing behavior problems</b>	Child	BPM <sup>a</sup> 4,018	-	-	RADS-2 <sup>f</sup> + SCARED <sup>g</sup> 266	YSR <sup>b</sup> 2,171		
	Mother	CBCL <sup>c</sup> 4,55	CBCL <sup>c</sup> 021,731	-	-	CBCL <sup>c</sup> 1,955		
	Father	CBCL <sup>c</sup> 3,259	CBCL <sup>c</sup> 14,626	-	-	-		
	Teacher	-	TRF <sup>d</sup> 12,389	-	-	TCP <sup>e</sup> 1,924		

<sup>a</sup>Brief Problem Monitor (BPM; Achenbach, 2011).

<sup>b</sup>Youth Self Report (YSR; Achenbach, 1991).

<sup>c</sup>Child Behavior Checklist (CBCL; Achenbach, 1991; Achenbach, 2001).

<sup>d</sup>Teacher Report Form (TRF; Achenbach, 2001).

<sup>e</sup>Teacher Checklist of Psychopathology (TCP); Vignette questionnaire on the basis of the Achenbach Teacher Report Form developed by TRAILS.

<sup>f</sup>Reynolds Adolescent Depression Scale – 2nd edition (RADS-2; Reynolds, 2000). Excluding anhedonia scale. Standardized before averaged with SCARED.

<sup>g</sup>Screen for Child Anxiety Related Disorders (SCARED; Birmaher, et al., 1997). Standardized before averaged with RADS-2.



**Table 4:** Mean and SD for Externalizing and Internalizing Problems

<b>Informant</b>	<b>Cohort</b>	<b>Externalizing</b>	<b>Internalizing</b>	<b><i>N</i>-Ext/<i>N</i>-Int</b>
<b>Child</b>	Gen-R	1.94 (1.92)	2.15 (2.09)	4,010/4,018
	RADAR-Y	10.61 (7.15)	-0.04 (0.86)	491/266
	TRAILS	8.68 (6.25)	11.28 (7.41)	2,188/2,171
<b>Mother</b>	Gen-R	3.92 (4.91)	4.86 (5.05)	4,549/4,550
	NTR	5.61 (6.12)	4.68 (5.07)	11,086/10,986
	TRAILS	8.40 (7.03)	7.85 (6.20)	1,965/1,955
<b>Father</b>	Gen-R	3.99 (4.91)	4.58 (4.72)	3,259/3,259
	NTR	4.66 (5.41)	3.56 (4.24)	7,420/7,374
<b>Teacher</b>	NTR	3.28 (5.88)	4.41 (4.96)	6,536/6,446
	TRAILS	0.44 (0.77)	0.99 (1.12)	1,925/1,924

For instruments, see Note Table 3.

and internalizing behavior problem score, as a function of informant (father, mother, teacher and self).

## Measures

**Predictors.** Maternal and Paternal Age at Birth. The age of the biological parents at birth of the child was measured in years up to two decimals for each cohort.

**Outcomes.** Externalizing and Internalizing Problems. In most cohorts, internalizing and externalizing problems were assessed by the parent-rated Child Behavior Checklist (CBCL; Achenbach, 1991; Achenbach & Rescorla, 2001),<sup>40,41</sup> the Youth Self-Report (YSR),<sup>40</sup> and the Teacher Report Form (TRF)<sup>41</sup>. These questionnaires contain a list of around 120 behavioral and emotional problems, which can be rated as 0 = *not true*, 1 = *somewhat or sometimes true*, or 2 = *very or often true in the past 6 months*. The broadband scale Internalizing problems includes the syndromes anxious/depressed behavior, withdrawn/depressed behavior, and somatic complaints; the broadband scale Externalizing problems involves aggressive and rule-breaking behavior. In TRAILS, the Teacher Checklist of Psychopathology (TCP) was developed to be completed by teachers. The TCP contains descriptions of problem behaviors corresponding to the syndromes of the TRF. Teachers rated the TCP on a 5-point scale.<sup>42</sup> In Gen-R, the YSR was replaced by the Brief Problem Monitor (BPM), containing six items for internalizing and seven items for externalizing behavior problems from the YSR. All items were scored on a 3-point scale. In RADAR-Y, internalizing behavior problems were assessed by a combined score of the Reynolds Adolescent Depression Scale-2nd edition (RADS-2)<sup>38</sup> and the Screen for Child Anxiety Related Emotional Disorders (SCARED)<sup>39</sup> questionnaires. The RADS-2 contained 23 items (the subscale anhedonia was deleted) and the SCARED contained 38 items, which were rated on a 4-point scale (1 = *almost never*, 2 = *hardly ever*, 3 = *sometimes*, 4 = *most of the time*) and 3-point scale (1 = *almost never*, 2 = *sometimes*, 3 = *often*), respectively.

Table 3 gives an overview of the rating instruments, the informants for each of the cohorts and the number of children in each cohort for each informant/instrument combination. A sum score was calculated per informant/instrument for the relevant items for externalizing and internalizing problems respectively. Table 4 shows the mean scores for externalizing and internalizing problems per cohort. The scores for girls and boys are given in Tables S1 and S2 of the supplementary materials, respectively.

**Covariates.** Socio-Economic Status (SES) and child gender. In Gen-R, SES was defined as a continuous variable (principal component) based on parental education and household income. In NTR, SES was a 5-level ordinal variable based on occupational level. In TRAILS, SES was a 3-level ordinal variable based on parental education, parental occupational status and household income. In RADAR-Y SES was a dichotomous variable based on parents' occupational level. Child gender was coded as male = 0 and female = 1.

## Missing Data and Data Imputation

*Missing Data.* For externalizing problem behavior, 15.9% of the child self-reports were missing for Gen-R, while for RADAR-Y and TRAILS these percentages were 1.2% and 1.9%, respectively. For mother reported data, 4.6% were missing for Gen-R, 13.7% for NTR and 11.9% for TRAILS. For father reported data, 31.7% were missing for Gen-R and 42.1% for NTR. For teacher reported data, 50.5% were missing for NTR and 13.7% for TRAILS. For internalizing problem behavior, the percentages were similar, except for child-reported data in RADAR-Y, where 46.4% was missing. For the predictor variables, age mother and age father, 0.3% and 1.3%, were missing for NTR, 0.0% and 14.4% for Gen-R, 0.4% and 9.7% for RADAR-Y, and 5.1% and 25.0% for TRAILS, respectively. For SES, the percentage of missing values was always below 3.0%, except for Gen-R where 22.3% was missing. For child gender, all cohorts had complete information.

Please note that the higher percentage for missing teacher- and father-reported data of NTR is due to the fact that NTR did not collect teacher-reported data at the initiation of the study and that NTR had not collected father-reported data in multiple birth years due to financial constraints. The higher percentage of missing self-reported data of internalizing problem behavior for RADAR-Y is caused by the fact that not all subscales on which the internalizing problem behavior score was based were collected from all participants.

*Data Imputation.* Missing data was handled by means of multiple imputation (Schafer & Graham, 2002; Van Buuren, 2012). When multiple imputation is used, the missing values are repeatedly (in this study 100 times) imputed, that is, replaced by values that are plausible given the child's scores that are not missing, resulting in 100, so-called, completed data sets. Subsequently, each completed data set is analyzed (for example, using a multiple regression) and the 100 analyses are summarized such that the fact that "artificial data" are created by imputation is properly accounted for. Multiple imputation proceeds along three steps:

1. *Determine which variables are to be used for imputation.* The variables used for imputation have to be chosen such that conditional on these variables the missing data are believed to be missing at random (MAR),<sup>43</sup> that is, whether or not a score is missing does not depend on the missing value.<sup>44</sup> Unless missingness is planned, the variables causing the missingness are unknown to the researcher. What is often done in practice is that variables are chosen that are expected to be good predictors of the variables containing missing values. One can argue with respect to which and how many variables to use, but there is no way to test whether MAR is achieved, and MAR is an assumption.

The imputation model included the outcome variables externalizing and internalizing behavioral problems per informant, total behavioral problems, SES, child gender, age of the child, age of the father and age of the mother. In some cohorts, other variables were present that could also contribute to the imputation. Specifically, parent



psychopathology (in Gen-R) and total number of siblings (in NTR) contributed to the imputation model. Variables functioned only as predictors when a correlation of at least .10 with the imputed variable was present. Since the NTR dataset contained twins, the imputation process differed from that of the other cohorts. The imputation for NTR was done for each family instead of each participant, so that the same value for SES, age father and age mother was obtained for both twins. The imputation of missing data was done for informants available in each cohort. So, for example, when a cohort had no teacher-reported data, teacher data were not imputed.

2. *Generate imputed data matrices.* The R package MICE (Multiple Imputation by Chained Equations)<sup>43</sup> was used to create 100 imputed data matrices. MICE uses an iterative procedure in which sequentially each variable is imputed conditional on the real and imputed values of the other variables. Continuous variables were imputed by predictive mean matching. Categorical variables were imputed using logistic regression.<sup>45</sup> Success of the imputation was evaluated by checking the events logged by the software, and by checking convergence plots for a lack of trends and proper mixing of the imputation chains.

3. *Analyze each imputed data set as desired and pool the results.* In the current study each of the 100 imputed data sets was analyzed using multiple regression or cluster linear regression. The results, for each regression coefficient, were 100 estimates and 100 standard errors of the estimate. As may be clear, each of the standard errors was too small because they are partly based on artificial imputed data. This was accounted for by properly pooling the results using Rubin's rules.<sup>43</sup> The variance over the 100 estimates reflects the uncertainty in the estimate due to missing values (in each of the 100 completed data sets different values are imputed). In Rubin's rules the variance of the 100 estimates is used to increase the standard errors such that they properly account for the fact that part of the data is imputed. Gen-R, TRAILS and RADAR-Y used the 'pool' function of MICE in R for summarizing the effects of the 100 separate imputed datasets, whereas NTR used the pooling option of Mplus instead of R, to appropriately take into account the family clustering of the twins in the same analysis. Both pooling methods are based on the principles as explained here. The pooled estimates and standard errors were the main outcomes of the analyses after imputation.

### **Analytical Strategy: Bayesian Evidence Synthesis**

The process of Bayesian evidence synthesis consists of four steps: (1) creating exploratory and confirmatory data sets; (2) generating competing hypotheses using exploratory analysis; (3) quantifying the support for each of the competing hypotheses using Bayesian hypothesis evaluation; and (4) Bayesian evidence synthesis, that is, sum-

marizing the support resulting from each study into the overall support for the competing hypotheses in the data from the four cohort studies.

### *Exploration and Confirmation*

As was elaborated in the introduction, diverse results regarding the relation between parental age and child problem behavior have been found in the literature, with increased parental age both positively and negatively related to child problem behavior. In the same vein, there may be a quadratic effect and if there is, the change in child problem behavior may be accelerating or decelerating across parental age. Since research is indecisive, especially for the non-clinical studies reviewed in this paper, the data resulting from each of the cohorts were split randomly into two parts containing the same number of children: an exploratory part, which was used to generate a set of competing hypotheses; and a confirmatory part, which was used to quantify the support in the data for each of the hypotheses considered. Since the NTR dataset consisted of twins, the cross-validation datasets were split based on family ID for this cohort, to ensure independent datasets. Multiple imputation was applied separately to the exploratory and confirmatory part of the data. Having an exploratory and confirmatory dataset avoids the so-called “double dipping”, that is, using *the same* data to generate and evaluate hypotheses. Here a hypothesis survived if it: 1) emerged from the exploratory analyses and 2) was supported by the confirmatory analyses. The process of generating hypotheses is explained below.

### *Generating Hypotheses using Exploratory Analyses*

The exploratory half of the data resulting from each of the four cohorts was used to generate hypotheses with respect to the relation between child problem behavior and parental age. First, for each cohort separately, linear regression analyses were conducted to regress internalizing and externalizing problem behavior as evaluated by child, mother, father, and teacher (See Table 3 for the informants that were present per cohort) on paternal and maternal age and age squared (both with and without child gender and social economic status as covariates). Parental age was mean-centered to obtain the linear effect at the mean age of the samples and to reduce the correlation between the linear and quadratic term. For Gen-R, RADAR-Y and TRAILS, the analyses were conducted in R (R Core Team, 2017). For the NTR twin-data, cluster linear regression analyses were conducted in Mplus version 8.0.<sup>46</sup> All analyses were repeated with SES and child gender as covariates. This rendered, for each combination (e.g., predicting externalizing problems as rated by the mother from mother age and age squared) an estimate of both the linear and quadratic effect for each of the cohorts that included the informant of interest. These estimates and the corresponding p-values provided information with respect to whether the linear and non-linear effects were expected to be negative, zero, or positive. To interpret the strength of relations, the variables in the exploratory analyses were all standardized. The results of the regression analyses were



translated into so-called informative hypotheses,<sup>47</sup> that is, hypotheses that represent expectations with respect to the state of affairs in the populations from which the data of the four cohorts were sampled. One example of such an informative hypothesis is:  $H1: \beta < 0$ . That is, the regression coefficient is negative. Informative hypotheses go beyond the traditional null hypothesis (here  $H0: \beta = 0$ ) by stating explicitly which relations between variables are expected. Often the null is added to the set of hypotheses under consideration to protect against unjustified claims that the effect specified by an informative hypothesis exists. Another hypothesis that can be added besides the informative hypotheses is the alternative hypothesis  $H_a: \beta$ . That is, there are no restrictions on the regression coefficient. The alternative hypothesis is used to protect against choosing the best of a set of inadequate informative hypotheses. For example,  $H0: \beta = 0$ , and  $H1: \beta < 0$  constitute the set of hypotheses supported by the exploratory parts of the data, but both are inadequate in the confirmatory data. Instead, another unspecified hypothesis ( $\beta > 0$ ) describes the confirmatory data best. In this case the Bayesian approach (specified below) will prefer the alternative hypothesis,  $H_a: \beta$ , over both informative hypotheses. By using informative hypotheses, the exact same hypotheses could be evaluated in all cohorts, even when cohorts used different measurement instruments for the same concepts. Not requiring the exact same measurement instruments is an important benefit of Bayesian evidence synthesis over classical meta-analyses.

### *Confirmatory Bayesian Hypotheses Evaluation*

Once a set of competing informative hypotheses had been formulated (including the traditional null and alternative hypotheses), the empirical support for each pair of hypotheses was quantified using the Bayes factor (BF).<sup>48</sup> The Bayes factor is the ratio of the marginal likelihood of two competing hypotheses. Loosely spoken, the marginal likelihood of a hypothesis is the probability of that hypothesis given the data. Consequently, a Bayes factor comparing  $H1$  with  $H_a$  of, for example, 5 indicates that the support in the data for  $H1$  is five times larger than for  $H_a$ . The BF as the ratio of two marginal likelihoods implies that the fit (how well does a hypothesis describe the data set at hand) and the specificity (how specific is a hypothesis of the hypotheses involved are accounted for).<sup>49</sup> To give an example, if  $\beta = -2$ ,  $H1: \beta < 0$ , and  $H_a: \beta$ , both have an excellent fit, but  $H1: \beta < 0$  is more specific than  $H_a: \beta$  (anything goes), and as a result, the BF will prefer  $H1$  over  $H_a$ . Note that the size of the Bayes factor is related to sample size. If the precision of the evidence in the data for a hypothesis increases as a result of a larger sample, the Bayes factor for that hypothesis will increase as well. The Bayes factor implemented in the R package *Bain*<sup>49</sup> was used to evaluate informative hypotheses in the context of (cluster) multiple linear regression models.

Assuming that a priori each hypothesis is equally likely to be true, the Bayes factors were transformed in so-called posterior model probabilities (PMPs), that is, the support in the data for the hypothesis at hand given the set of hypotheses under evaluation. PMPs have values between 0 and 1 and sum to 1 for the hypotheses in the set under



consideration. For example, if PMP H0 = .05, PMP H1 = .85, and PMP Ha = .10, then it is clear that H1 receives the most support from the data, because it has by far the largest PMP. Thus, the result of the confirmatory Bayesian hypotheses evaluation were PMPs for each hypothesis and for each informant by each of the cohorts that had ratings by this informant. The next step was to apply Bayesian evidence synthesis.

### *Bayesian Evidence Synthesis*

Bayesian evidence synthesis was used to summarize the support for the hypotheses of interest over the four cohort studies. Bayesian evidence synthesis<sup>37</sup> can be illustrated using the set of hypotheses: H0:  $\beta = 0$ , H1:  $\beta < 0$ , and Ha:  $\beta$ . In the context of this paper, these hypotheses are incompletely specified. The complete specification would be H0:  $\beta_1 = 0$  for NTR, H1:  $\beta_1 < 0$  for NTR and Ha:  $\beta_1$  for NTR, and analogously for the other three cohort studies. This specification highlights that the support for the hypotheses depends on the cohort study at hand. Bayesian evidence synthesis can then be used to determine support for a set of hypotheses:

H0: H0 for NTR & H0 for TRAILS & H0 for Gen-R & H0 for Radar-Y

H1: H1 for NTR & H1 for TRAILS & H1 for Gen-R & H1 for Radar-Y

Ha: Ha for NTR & Ha for TRAILS & Ha for Gen-R & Ha for Radar-Y

that is, the regression coefficient is zero in the populations corresponding to each of the four cohort studies, the regression coefficient is smaller than zero in the populations corresponding to each of the four cohort studies, and there is not prediction with respect to the regression coefficient in the populations corresponding to each of the four cohort studies. If for a specific set of hypotheses only two or three cohorts contain the necessary variables, the hypotheses can be adjusted accordingly. Like for each individual study, the support for these composite hypotheses was quantified using posterior model probabilities (PMPs).

If a hypothesis emerges from the exploratory analyses of the data corresponding to the cohort studies and is supported by the confirmatory analyses of the data corresponding to the cohort studies, then there is evidence that this hypothesis provides an adequate description of the relation between child problem behavior and parental age, that is, in general, independent of the specific cohort studies used to evaluate this hypothesis. With the methodological approach elaborated in this section and applied in the remainder of this paper, the increased awareness of the need for replication studies before making scientific claims is explicitly addressed.

## **RESULTS**



## Exploratory Analyses

The results of the exploratory analyses (see Supplementary Materials) generally showed a negative relation between mean-centered parental age and externalizing problems accompanied by a positive quadratic coefficient, implying that the negative relation with age at the mean declined across age (see Table S3 of the Supplemental Materials). This model explained about 1.9% of total variance in externalizing problems with maternal age and 1.2% with paternal age. For internalizing problems, the relation with parental age was less apparent: about 0.5% of the total variance was explained by mothers' age, and about 0.2% was explained by fathers' age. In analyses including the covariates SES and gender, the relation with age diminished, but remained significant (Tables S4, and S5, of the supplementary materials). Higher SES was related to fewer externalizing problems, and boys showed more externalizing problems than girls. In general, no relation between parental age and internalizing problems was observed (see Tables S6, S7, and S8 of the Supplemental Materials).

Our interpretation of the exploratory results led to the following set of competing informative hypotheses with respect to the relation between parental age (mean-centered), as indicated by a linear (i.e.,  $\beta_1$ ) and quadratic (i.e.,  $\beta_2$ ) coefficient, and child problem behavior:

- H1:  $\beta_1 = 0$ ,  $\beta_2 = 0$ . That is, age does not have a linear or quadratic relation.
- H2:  $\beta_1 < 0$ ,  $\beta_2 = 0$ . That is, age has a negative linear relation, there is no quadratic relation.
- H3:  $\beta_1 < 0$ ,  $\beta_2 > 0$ . That is, age has a negative linear relation, and a positive quadratic relation.
- Ha:  $\beta_1$ ,  $\beta_2$ . That is, none of the above.

Based on the exploratory results, we expected most evidence for H2 or H3 in analyses with parental age predicting externalizing problems, and most evidence for H1 in analyses with parental age predicting internalizing problems. Since the exploratory results did not show a positive linear or a negative quadratic relation between age and behavioral problems, the hypotheses do not include these features. However, we remained open to other options by including the alternative hypothesis Ha that imposes no constraints on the parameters, and accordingly claims that anything can be true. Ha receives the most support if none of the specified informative hypotheses provides an adequate description of the confirmatory part of the data from each of the four cohorts. In this manner, we avoided that the best hypothesis out of the set of H1, H2, and H3, is an implausible hypothesis.

## Confirmatory Analyses

Similarly to the exploratory data, the results showed negative relations across cohorts between parental age and externalizing problems. However, in the confirmatory

**Table 5: Posterior Model Probabilities for Parental Age Predicting Externalizing Problems**

Informant	Cohort	Age Father				Age Mother			
		H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>
<b>Child</b>	Gen-R	.23	.56	.16	.05	.22	.18	.49	.13
	RADAR-Y	.28	.02	.49	.22	.43	.07	.38	.12
	TRAILS	.86	.13	.00	.01	.83	.15	.02	.01
	All	<b>.98</b>	.02	.00	.00	<b>.93</b>	.02	.04	.00
<b>Mother</b>	Gen-R	.90	.07	.02	.01	.82	.04	.10	.05
	NTR	.00	.02	.74	.24	.00	.89	.09	.03
	TRAILS	.18	.74	.06	.02	.00	.88	.09	.03
	All	.00	<b>.53</b>	<b>.45</b>	.00	.00	<b>.97</b>	.03	.00
<b>Father</b>	Gen-R	.65	.22	.10	.03	.60	.19	.17	.04
	NTR	.00	.49	.38	.13	.00	.93	.05	.02
	All	.00	<b>.73</b>	.25	.02	.00	<b>.95</b>	.05	.00
<b>Teacher</b>	NTR	.55	.41	.03	.01	.29	.60	.09	.02
	TRAILS	.48	.31	.16	.05	.00	.73	.21	.06
	All	<b>.67</b>	.32	.01	.00	.00	<b>.96</b>	.04	.00

Numbers in italic font represent the highest posterior model probability per cohort. Numbers in bold font represent the highest meta-analytic results

**Table 6:** Posterior Model Probabilities for Parental Age Predicting Externalizing Problems after Correction for Impact Covariates

Informant	Cohort	Age Father				Age Mother			
		H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>
<b>Child</b>	Gen-R	.62	.33	.04	.01	.83	.10	.05	.02
	RADAR-Y	.36	.02	.42	.19	.53	.08	.29	.10
	TRAILS	.88	.11	.00	.01	.89	.09	.02	.01
	All	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00
<b>Mother</b>	Gen-R	.96	.03	.00	.00	.97	.02	.00	.01
	NTR	.00	.31	.52	.17	.00	.95	.04	.01
	TRAILS	.67	.31	.01	.01	.30	.63	.05	.02
	All	.03	<b>.99</b>	.00	.00	.00	<b>1.00</b>	.00	.00
<b>Father</b>	Gen-R	.88	.10	.02	.00	.92	.06	.01	.00
	NTR	.02	.84	.11	.04	.00	.96	.03	.01
	All	.15	<b>.84</b>	.02	.00	.00	<b>.99</b>	.01	.00
<b>Teacher</b>	NTR	.79	.20	.01	.00	.68	.28	.03	.01
	TRAILS	.87	.11	.02	.00	.60	.32	.07	.02
	All	<b>.97</b>	.03	.00	.00	<b>.81</b>	.18	.00	.00



**Table 7: Posterior Model Probabilities for Parental Age Predicting Internalizing Problems**

Informant	Cohort	Age Father					Age Mother				
		H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>		
<b>Child</b>	Gen-R	.91	.08	.01	.00	.86	.09	.04	.01	.01	
	RADAR-Y	.84	.09	.05	.03	.81	.16	.02	.01	.01	
	TRAILS	.96	.04	.00	.00	.93	.06	.01	.00	.00	
	All	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00	.00	
<b>Mother</b>	Gen-R	.58	.25	.14	.04	.35	.25	.33	.08	.08	
	NTR	.69	.26	.04	.01	.26	.72	.01	.01	.01	
	TRAILS	.94	.05	.00	.00	.81	.17	.02	.01	.01	
	All	<b>.99</b>	.01	.00	.00	<b>.71</b>	.29	.00	.00	.00	
<b>Father</b>	Gen-R	.43	.42	.11	.03	.48	.36	.13	.03	.03	
	NTR	.96	.04	.00	.00	.95	.05	.00	.00	.00	
	All	<b>.96</b>	.04	.00	.00	<b>.97</b>	.03	.00	.00	.00	
<b>Teacher</b>	NTR	.99	.01	.1	.00	.99	.01	.00	.00	.00	
	TRAILS	.85	.06	.07	.02	.24	.15	.49	.12	.12	
	All	<b>1.00</b>	.00	.00	.00	<b>.99</b>	.01	.00	.00	.00	

**Table 8:** Posterior Model Probabilities for Parental Age Predicting Internalizing Problems after Correction for Impact Covariates

Informant	Cohort	Age Father					Age Mother						
		H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>
<b>Child</b>	Gen-R	.77	.21	.02	.01	.82	.09	.07	.02	.82	.09	.07	.02
	RADAR-Y	.86	.07	.04	.03	.86	.11	.02	.01	.86	.11	.02	.01
	TRAILS	.97	.03	.00	.00	.95	.04	.00	.00	.95	.04	.00	.00
	All	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00
<b>Mother</b>	Gen-R	.88	.11	.01	.00	.93	.05	.01	.00	.93	.05	.01	.00
	NTR	.88	.11	.01	.00	.70	.29	.00	.00	.70	.29	.00	.00
	TRAILS	.96	.04	.00	.00	.91	.08	.01	.00	.91	.08	.01	.00
	All	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00
<b>Father</b>	Gen-R	.88	.09	.02	.01	.90	.08	.01	.00	.90	.08	.01	.00
	NTR	.96	.03	.00	.00	.96	.04	.00	.00	.96	.04	.00	.00
	All	<b>1.00</b>	.01	.00	.00	<b>1.00</b>	.01	.00	.00	<b>1.00</b>	.01	.00	.00
	NTR	.99	.01	.00	.00	.99	.01	.00	.00	.99	.01	.00	.00
<b>Teacher</b>	TRAILS	.94	.04	.02	.01	.83	.06	.08	.01	.83	.06	.08	.03
	All	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	t.00	.00	<b>1.00</b>	.00	t.00	.00

Numbers in italic font represent the highest posterior model probability per cohort. Numbers in **bold** font represent the highest meta-analytic results.



**Table 7:** Posterior Model Probabilities for Parental Age Predicting Internalizing Problems

Informant	Cohort	Age Father				Age Mother			
		H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>
Child	Gen-R	.91	.08	.01	.00	.86	.09	.04	.01
	RADAR-Y	.84	.09	.05	.03	.81	.16	.02	.01
	TRAILS	.96	.04	.00	.00	.93	.06	.01	.00
	<i>All</i>	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00
Mother	Gen-R	.58	.25	.14	.04	.35	.25	.33	.08
	NTR	.69	.26	.04	.01	.26	.72	.01	.01
	TRAILS	.94	.05	.00	.00	.81	.17	.02	.01
	<i>All</i>	<b>.99</b>	.01	.00	.00	<b>.71</b>	.29	.00	.00
Father	Gen-R	.43	.42	.11	.03	.48	.36	.13	.03
	NTR	.96	.04	.00	.00	.95	.05	.00	.00
	<i>All</i>	<b>.96</b>	.04	.00	.00	<b>.97</b>	.03	.00	.00
	Teacher	NTR	.99	.01	.1	.00	.99	.01	.00
Teacher	TRAILS	.85	.06	.07	.02	.24	.15	.49	.12
	<i>All</i>	<b>1.00</b>	.00	.00	.00	<b>.99</b>	.01	.00	.00

Numbers in **bold** font represent the posterior model probability per cohort. Numbers in *italic* font represent the meta-analytic results.

data, the quadratic coefficients from the cohorts were less often significantly different from zero than in the exploratory data. The model with a linear and quadratic coefficient for parental age explained on average about 1.1% of total variance in externalizing problems with maternal age and 0.9% with paternal age as a predictor. With respect to internalizing behavior problems, the model with maternal age explained on average about 0.4% of the total variance, and paternal age explained on average about 0.3%.

### *Parental Age and Externalizing Behavior Problems*

The posterior model probabilities (PMPs) concerning the relation between parental age and externalizing problems are presented in Tables 5. The table only shows PMP scores for those cohorts that included the associated informants (See Table 3 for an overview of informants per cohort). As shown by Table 5, for parent-reported externalizing behavior problems, Gen-R yielded most evidence for H1 (i.e., no relation with parental age). NTR supports H3 (i.e., the relation with parental age follows a negative linear trend including a positive quadratic factor) for mother-reported externalizing behavior problems, while TRAILS provided most support for H2 (i.e., the relation with parental age is linear and negative). The combined results for mother-reported externalizing behavior problems predicted by father age showed substantial support (PMP = .53 and .45 respectively) for H2 and H3. For father reported externalizing behavior

**Table 8:** Posterior Model Probabilities for Parental Age Predicting Internalizing Problems after Correction for Impact Covariates

Informant	Cohort	Age Father				Age Mother			
		H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>a</sub>
Child	Gen-R	.77	.21	.02	.01	.82	.09	.07	.02
	RADAR-Y	.86	.07	.04	.03	.86	.11	.02	.01
	TRAILS	.97	.03	.00	.00	.95	.04	.00	.00
	<i>All</i>	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00
Mother	Gen-R	.88	.11	.01	.00	.93	.05	.01	.00
	NTR	.88	.11	.01	.00	.70	.29	.00	.00
	TRAILS	.96	.04	.00	.00	.91	.08	.01	.00
	<i>All</i>	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00
Father	Gen-R	.88	.09	.02	.01	.90	.08	.01	.00
	NTR	.96	.03	.00	.00	.96	.04	.00	.00
	<i>All</i>	<b>1.00</b>	.01	.00	.00	<b>1.00</b>	.01	.00	.00
Teacher	NTR	.99	.01	.00	.00	.99	.01	.00	.00
	TRAILS	.94	.04	.02	.01	.83	.06	.08	.03
	<i>All</i>	<b>1.00</b>	.00	.00	.00	<b>1.00</b>	.00	.00	.00

Numbers in **bold** font represent the posterior model probability per cohort. Numbers in *italic* font represent the meta-analytic results.

problems predicted by father age and for parent-reported externalizing behavior problems predicted by mother age, the combined results provided most support for H2: the relation with parental age is linear and negative, in other words, higher parental age is associated with less externalizing behavioral problems. For teacher-reported externalizing behavior problems, TRAILS and NTR yielded most evidence for H1 when paternal age was used as a predictor in the linear regression model. When maternal age was included, most support was yielded for H2: the relation with parental age is linear and negative. For child-reported externalizing behavior problems, the results were mixed over cohorts (Gen-R prefers H2, RADAR-Y H3, and TRAILS H1). After combining the results from the three cohorts, however, most support was obtained for H1, that is, no relation with parental age.

Table 6 shows the results after inclusion of the covariates as predictors of externalizing problems. After adjusting for the effect of SES and gender, all cohorts yielded substantial evidence for H1 with respect to child- and teacher-reported externalizing problem behavior. This meant a shift especially for the child-reported problem behavior by Gen-R, and the teacher-reported problem behaviors by both NTR and TRAILS. For



parent reported problem behavior some cohorts provided most support for H1 (Gen-R for all parent-reports, and TRAILS for paternal age predicting mother-reported problem behavior), others for H2 (TRAILS and NTR), and NTR for H3 in mother-reported problem scores related to paternal age. By including covariates in the model Gen-R and TRAILS mainly handed in support on H2 while in NTR the support for H2 increased at the expense of support for H3. When combining evidence for the parent reports most support was still found for H2, that is, there is a linearly decreasing relation between age and externalizing problem behavior.

### *Parental Age and Internalizing Behavior Problems*

With regard to internalizing problems (the results are presented in Table 7), the cohorts generally found most evidence for H1 for multiple informants, except for mother-reported internalizing problems reported by mother age in NTR. All combinations of studies rendered most support for H1, which means that the hypothesis that there is no relation between parental age and internalizing problems is best supported by the data.

After correction for SES and gender (Table 8), all findings still suggested H1 for the impact of parental age on internalizing problem behavior, irrespective of the cohort and informant, and, consequently, combining the results from the various cohorts provided overwhelming support for H1, that is, no evidence for a relation between parental age and child internalizing problem behavior.

## **DISCUSSION**

### **Parental Age and Externalizing Problems**

We found evidence for a negative linear relation between parental age and externalizing problems as reported by parents. That is, across increasing maternal and paternal age there was a decrease in offspring externalizing problems. There was also evidence for a negative linear relation between maternal age and externalizing problems as reported by teachers. For teachers, this finding was partly explained by socio-economic status, but the relation between parental age and parent-reported externalizing problems persisted after adjusting for SES, so the favorable effect of parental age is not solely due to socio-economic status.

### **Parental Age and Internalizing Problems**

Parental age seemed unrelated to child internalizing problem behavior, especially when taking SES into account. Tentatively, older parenthood might be associated with both high and low vulnerability to develop internalizing problems. On the one hand, older parents may have a lower probability of internalizing problems because they are less likely to have a background characterized by deprivation and social instability,<sup>50</sup> known to be related to internalizing problems such as anxiety and depression. On the

other hand, internalizing problems can increase the probability of older parenthood, by stimulating engagement in and consolidation of romantic relationships.<sup>51,52</sup> Possibly, both processes play a role, and their joint influence results in a lack of net result.

### **Sociodemographic Factors as a Potential Explanation**

The relatively consistent beneficial effect of advanced parenthood for childhood externalizing problems may seem unexpected, given mixed findings from earlier research on more common mental health problems (De Kluiver, Buizer-Voskamp, Dolan, & Boomsma, 2017; McGrath et al., 2014).<sup>11,12</sup> The beneficial effect of advanced parental age could have more than one explanation. Older and younger parents have different parenting styles. For example, there is evidence that older mothers use less frequent sanctions towards their children, are more sensitive to the child's needs and provide more structure.<sup>53</sup> Older parents may also tend to appraise a specific problem level as less disturbing than younger parents, and older parents might be more patient and are capable of setting limits, thus feeling more equipped to handle externalizing behaviors. The positive impact of higher quality parenting by older parents is expected to be more relevant to externalizing problem behavior than to autism and schizophrenia where a disadvantageous impact of increased parental age is established.

Previous studies provided evidence indicating that offspring of older parents are, in several respects, more affluent than those with younger parents.<sup>3,7,11,21,22,25</sup> The fact that the negative relation of parental age and externalizing problems became weaker when SES was taken into account, indicates that the relatively high socio-economic status of older parents, or SES-related selection effects,<sup>50</sup> at least partly explain why their children have a decreased probability of externalizing problems. Myrskylä, Barclay and Goisis<sup>24</sup> argued that there are indeed important socio-demographic pathways associated with delayed parenthood in more recent birth cohorts. Older mothers tend to have better health behaviors during pregnancy, for example with respect to smoking during pregnancy, which is an established risk factor for offspring externalizing problems.<sup>54</sup>

Furthermore, parents who have externalizing behavior problems themselves may be higher in risk taking and may have children at a younger age. Hence, externalizing behavior problems may be transmitted especially by younger parents and less by older parents. This idea is in line with the unclarity about a relation between ADHD and advanced paternal age.<sup>11,12</sup>

From a biological point of view, advanced parenthood seems mostly disadvantageous, but sociodemographic factors might compensate or even more than compensate for the biological disadvantages related to reproductive ageing when it comes to mental health problems. Older mothers from more recent birth cohorts are more socioeconomically advantaged, and happier after childbearing. The observation that older parents have offspring with fewer externalising problems, tended to disappear when

SES was taken into account, shows that demographic factors can indeed compensate for the biological disadvantages.

### **Earlier Versus Later Birth Cohorts**

In the 1950's and 60's the number children born to mothers over 40 was larger than in 2016. For offspring born during the 1960s, Saha et al.<sup>55</sup> found a negative association between maternal age and externalizing behavior problems, but in contrast to our results, they observed a positive association between maternal age and internalizing problems, and a positive association between paternal age and externalizing behavior problems. The study differed in several important aspects from the current one. All offspring were born during the 1960s, whereas in our study, all offspring were born after 1980. The age at which fathers and mothers have children has increased in the last 20 years. In the Saha et al. study average maternal and paternal ages were 24.8 and 28.4, respectively, while in our samples average maternal- and paternal ages were around 31 and 33 years. Older mothers from earlier birth cohorts tended to have low levels of education and their offspring had many older siblings.<sup>24</sup> In later birth cohorts, older mothers have higher education than younger mothers and their offspring have fewer older siblings. Thus, the family resources are spread less thinly across siblings than in earlier times. This may be the reason that our results differ from some of the findings of Saha, Barnett, Buka and McGrath.<sup>55</sup> As argued by Myrskylä, Barclay & Goisis,<sup>24</sup> as well, being a parent during the 1960s differs from being a parent in the 1980s and children born during the 1980s and later might benefit from positive changes in the macro-environment.

### **Informant Effect**

We used a multi-informant design (i.e., mother, father, teacher, child) to investigate parental age effects on behavioral problems. Most questionnaires belonged to the same system (ASEBA), but they do not necessarily capture the exact same construct, as different informants observe the children in different contexts. Consistent with the notion that different informants provide partly non-overlapping information, the results in this study depended on the choice of informant, since, as opposed to parent-reported problems, child-reported externalizing problems were not predicted by parental age. Conceivably, this different outcome for child-reported problems is due to a limited ability of 10-year-old children to report reliably and validly on their externalizing behaviors. It is less likely that the associations with parent-reports are caused by report bias, as teacher-reports also provide support for an association with maternal age. Thus, the choice of informant is not an arbitrary one, and may influence the associations that are found. Obviously, the parent and teacher sample sizes were also substantially larger than the sample size for child-reports. Additionally, the largest study with child reports

(i.e., TRAILS) used a shortened version of the YSR, which could cause lower reliability and validity of child-reports.

### **Strengths of the Current Study**

This study adopted an analysis strategy that used the data of multiple cohort studies to evaluate the same set of hypotheses. First, the data of each cohort study were divided into two parts: an exploratory part and a confirmatory part. Second, the exploratory part was used to generate a set of competing informative hypotheses. Third, the confirmatory part was used to compute the support in each cohort for the hypotheses entertained and to combine studies by means of Bayesian updating to compute overall results.<sup>37</sup> This analysis strategy had a number of advantages. In the exploratory analyses data snooping or even p-hacking is allowed, because this part of the data is only used to generate a set of competing informative hypotheses and not to evaluate these hypotheses. In contrast, the confirmatory part of each data set is only used to evaluate this set of informative hypotheses to the traditional null and alternative hypotheses, which should, especially in ages of replication crisis, publication bias and questionable research practices, increase the credibility of our results. The interested reader is referred to the Supplementary Materials where we highlight, why exploratory analyses may lead to incorrect interpretations, even with large samples, and that cross-validation can prevent this from happening. In addition, with traditional null hypothesis significance testing, we would not have been able to quantify the support for the null hypothesis (p-values cannot be used to “accept” the null-hypothesis), which appeared an important hypothesis in our study. Bayes factors and posterior model probabilities are not used to reject or not reject the null-hypotheses, they are used to quantify the support in each of the cohorts for the hypotheses entertained. Furthermore, combining studies using Bayesian updating enabled us to quantify the relative evidence with respect to multiple hypotheses using the data from multiple cohorts. Again, in ages of replication crisis, it is valuable to base conclusions on data from multiple cohorts that can all be used to address the same research question.

### **Limitations**

Although the study has a number of methodological strengths, there are also limitations. First, the study focused on children’s externalizing and internalizing behavior problems and did not examine other outcomes that may be positively associated with parental age, such as physical health problems and neurodevelopmental conditions. Second, children’s behavior problems were only assessed during early adolescence. Thus, the study could not investigate the possibility that the direction or magnitude of the associations may vary at different points in development. For example, previous research suggesting a negative association between parental age and individuals’ well-being has focused on late adolescents and young adults.<sup>7,8</sup> Third, a tiny percentage of the parents were under the age of 20 at the time of the child’s birth. Although this

reflects societal changes in Netherlands, it would be important to note that some results may not replicate in other populations that have a higher percentage of teenage pregnancies. This may be especially relevant when interpreting the lack of an association between parental age and children's internalizing behavior problems in this study.

### **Conclusion**

The strategy applied to large cohorts showed us a beneficial association between advanced parental age and externalizing problem behavior, while for internalizing problem behavior there is no beneficial association. We found no evidence for a harmful effect of advanced parenthood.

## REFERENCES

1. te Velde ER. The variability of female reproductive ageing. *Hum Reprod Update*. 2002;8(2):141-154. doi:10.1093/humupd/8.2.141
2. Desjardins B, Bideau A, Brunet G. Age of mother at last birth in two historical populations. *J Biosoc Sci*. 1994;26(04). doi:10.1017/s0021932000021635
3. Carslake D, Tynelius P, van den Berg G, Smith GD, Rasmussen F. Associations of parental age with health and social factors in adult offspring. Methodological pitfalls and possibilities. *Sci Rep*. 2017;7(1). doi:10.1038/srep45278
4. Sandin S, Hultman CM, Kolevzon A, Gross R, MacCabe JH, Reichenberg A. Advancing Maternal Age Is Associated With Increasing Risk for Autism: A Review and Meta-Analysis. *J Am Acad Child Adolesc Psychiatry*. 2012;51(5):477-486.e1. doi:10.1016/j.jaac.2012.02.018
5. Lee BK, McGrath JJ. Advancing parental age and autism: multifactorial pathways. *Trends Mol Med*. 2015;21(2):118-125. doi:10.1016/j.molmed.2014.11.005
6. Menezes PR, Lewis G, Rasmussen F, et al. Paternal and maternal ages at conception and risk of bipolar affective disorder in their offspring. *Psychol Med*. 2010;40(03):477. doi:10.1017/s003329170999064x
7. Tearne JE, Robinson M, Jacoby P, et al. Older maternal age is associated with depression, anxiety, and stress symptoms in young adult female offspring. *J Abnorm Psychol*. 2016;125(1):1-10. doi:10.1037/abn0000119
8. Weiser M, Reichenberg A, Werbeloff N, et al. Advanced Parental Age at Birth Is Associated With Poorer Social Functioning in Adolescent Males: Shedding Light on a Core Symptom of Schizophrenia and Autism. *Schizophr Bull*. 2008;34(6):1042-1046. doi:10.1093/schbul/sbn109
9. Andersen A-MN, Urhoj SK. Is advanced paternal age a health risk for the offspring? *Fertil Steril*. 2017;107(2):312-318. doi:10.1016/j.fertnstert.2016.12.019
10. Khandwala YS, Baker VL, Shaw GM, Stevenson DK, Lu Y, Eisenberg ML. Association of paternal age with perinatal outcomes between 2007 and 2016 in the United States: population based cohort study. *BMJ*. 2018;k4372. doi:10.1136/bmj.k4372
11. McGrath JJ, Petersen L, Agerbo E, Mors O, Mortensen PB, Pedersen CB. A Comprehensive Assessment of Parental Age and Psychiatric Disorders. *{JAMA} Psychiatry*. 2014;71(3):301. doi:10.1001/jamapsychiatry.2013.4081
12. de Kluiver H, Buizer-Voskamp JE, Dolan C V, Boomsma DI. Paternal age and psychiatric disorders: A review. *Am J Med Genet Part B Neuropsychiatr Genet*. 2017;174(3):202-213. doi:10.1002/ajmg.b.32508
13. Goisis A, Schneider DC, Myrskylä M. The reversing association between advanced maternal age and child cognitive ability: evidence from three {UK} birth cohorts. *Int J Epidemiol*. 2017;46(3):850-859. doi:10.1093/ije/dyw354
14. Jónsson H, Sulem P, Kehr B, et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*. 2017;549(7673):519-522. doi:10.1038/nature24018
15. Kong A, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. *Nature*. 2012;488(7412):471-475. doi:10.1038/nature11396
16. Francioli LC, and Paz P. Polak, Koren A, et al. Genome-wide patterns and properties of de novo mutations in humans. *Nat Genet*. 2015;47(7):822-826. doi:10.1038/ng.3292

17. Goldmann JM, Seplyarskiy VB, Wong WSW, et al. Germline De Novo Mutation Clusters Arise During Oocyte Aging in Genomic Regions With High Double-Strand-Break Incidence. *Obstet Gynecol Surv.* 2018;73(9):531-532. doi:10.1097/ogx.0000000000000604
18. Wong WSW, Solomon BD, Bodian DL, et al. New observations on maternal age effect on germline de novo mutations. *Nat Commun.* 2016;7(1). doi:10.1038/ncomms10486
19. Jenkins TG, Aston KI, Pflueger C, Cairns BR, Carrell DT. Age-Associated Sperm {DNA} Methylation Alterations: Possible Implications in Offspring Disease Susceptibility. Greally JM, ed. {PLoS} Genet. 2014;10(7):e1004458. doi:10.1371/journal.pgen.1004458
20. Gratten J, Wray NR, Peyrot WJ, McGrath JJ, Visscher PM, Goddard ME. Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations. *Nat Genet.* 2016;48(7):718-724. doi:10.1038/ng.3577
21. Myrskylä M, Fenelon A. Maternal Age and Offspring Adult Health: Evidence From the Health and Retirement Study. *Demography.* 2012;49(4):1231-1257. doi:10.1007/s13524-012-0132-x
22. Tearne JE, Robinson M, Jacoby P, Li J, Newnham J, McLean N. Does Late Childbearing Increase the Risk for Behavioural Problems in Children? A Longitudinal Cohort Study. *Paediatr Perinat Epidemiol.* 2015;29(1):41-49. doi:10.1111/ppe.12165
23. Janecka M, Haworth CMA, Ronald A, et al. Paternal Age Alters Social Development in Offspring. *J Am Acad Child Adolesc Psychiatry.* 2017;56(5):383-390. doi:10.1016/j.jaac.2017.02.006
24. Myrskylä M, Barclay K, Goisis A. Advantages of later motherhood. GynaOrlebeke JF, Knol DL, Boomsma DI, Verhulst FC. Frequency of Parental Report of Problem Behavior in Children Decreases with Increasing Maternal Age at Delivery. *Psychol Rep.* 1998;82(2):395-404. doi:10.2466/pr0.1998.82.2.395
25. Orlebeke JF, Knol DL, Boomsma DI, Verhulst FC. Frequency of Parental Report of Problem Behavior in Children Decreases with Increasing Maternal Age at Delivery. *Psychol Rep.* 1998;82(2):395-404. doi:10.2466/pr0.1998.82.2.395
26. Lawlor DA, Mortensen L, Andersen A-MN. Mechanisms underlying the associations of maternal age with adverse perinatal outcomes: a sibling study of 264 695 Danish women and their firstborn offspring. *Int J Epidemiol.* 2011;40(5):1205-1214. doi:10.1093/ije/dyr084
27. Bray I, Gunnell D, Smith GD. Advanced paternal age: How old is too old? *J Epidemiol Community Heal.* 2006;60(10):851-853. doi:10.1136/jech.2005.045179
28. McMahon CA, Gibson FL, Allen JL, Saunders D. Psychosocial adjustment during pregnancy for older couples conceiving through assisted reproductive technology. *Hum Reprod.* 2007;22(4):1168-1174. doi:10.1093/humrep/del502
29. Kiernan KE. Becoming a Young Parent: A Longitudinal Study of Associated Factors. *Br J Sociol.* 1997;48(3):406. doi:10.2307/591138
30. Janecka M, Rijdsdijk F, Rai D, Modabbernia A, Reichenberg A. Advantageous developmental outcomes of advancing paternal age. *Transl Psychiatry.* 2017;7(6):e1156. doi:10.1038/tp.2017.125
31. Crockett LJ, Bingham CR, Chopak JS, Vicary JR. Timing of first sexual intercourse: The role of social control, social learning, and problem behavior. *J Youth Adolesc.* 1996;25(1):89-111. doi:10.1007/bf01537382
32. Bailey JA, Hill KG, Oesterle S, Hawkins JD. Parenting practices and problem behavior across three generations: Monitoring, harsh discipline, and drug use in the intergenerational transmission of externalizing behavior. *Dev Psychol.* 2009;45(5):1214-1226. doi:10.1037/a0016129
33. Centraal Bureau voor de Statistiek. Geboorte; kerncijfers. 2018. <https://opendata.cbs.nl/stat->

line/#/CBS/nl/dataset/37422ned/table?ts=1522410899684.

34. Rescorla LA, Ginzburg S, Achenbach TM, et al. Cross-Informant Agreement Between Parent-Reported and Adolescent Self-Reported Problems in 25 Societies. *J Clin Child Adolesc Psychol.* 2013;42(2):262-273. doi:10.1080/15374416.2012.717870
35. Hudziak JJ, van Beijsterveldt CEM, Bartels M, et al. Individual differences in aggression: Genetic analyses by age, gender, and informant in 3-, 7-, and 10-year-old Dutch twins. *Behav Genet.* 2003;33(5):575-589. doi:10.1023/a:1025782918793
36. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* (80- ). 2015;349(6251). doi:10.1126/science.aac4716
37. Kuiper RM, Buskens V, Raub W, Hoijtink H. Combining Statistical Evidence From Several Studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociol Methods Res.* 2012;42(1):60-81. doi:10.1177/0049124112464867
38. Reynolds WM. Reynolds Adolescent Depression Scale. 2000. doi:10.1002/9780470479216.corpsy0798
39. Birmaher B, Khetarpal S, Brent D, et al. The screen for child anxiety related emotional disorders (SCARED): Scale construction and psychometric characteristics. *J Am Acad Child Adolesc Psychiatry.* 1997;36(4):545-553. doi:10.1097/00004583-199704000-00018
40. Achenbach TM, Edelbrock C. Child behavior checklist. 1991:546-552.
41. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms and Profiles. . (University of Vermont Youth, and Families RC for C, ed.). Burlington, VT; 2001.
42. de Winter AF, Oldehinkel AJ, Veenstra R, Brunnekreef JA, Verhulst FC, Ormel J. Evaluation of non-response bias in mental health determinants and outcomes in a large sample of pre-adolescents. *Eur J Epidemiol.* 2005;20(2):173-181. doi:10.1007/s10654-004-4948-6
43. Van Buuren S. Flexible Imputation of Missing Data. Chapman & Hall/CRC; 2012. doi:10.1201/b11826
44. Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychol Methods.* 2002;7(2):147-177. doi:10.1037//1082-989x.7.2.147
45. Van Buuren S, Groothuis-Oudshoorn K. {MICE}: Multivariate imputation by chained equations in {R}. *J Stat Softw.* 2011;45(3):1-67. doi:10.18637/jss.v045.i03
46. Muthén LK, Muthén BO. How to use a {M}onte {C}arlo study to decide on sample size and determine power. *Struct Equ Model.* 2002;9:599-620. doi:10.1207/S15328007SEM0904\_8
47. Hoijtink H. Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists. CRC Press; 2012. doi:10.1201/b11158
48. Kass RE, Raftery AE. Bayes Factors. *J Am Stat Assoc.* 1995;90(430):773-795. doi:10.1080/01621459.1995.10476572
49. Gu X, Mulder J, Hoijtink H. Approximated adjusted fractional {B}ayes factors: A general method for testing informative hypotheses. *Br J Math Stat Psychol.* 2018;71(2):229-261. doi:10.1111/bmsp.12110
50. Robson K, Pevalin DJ. Gender differences in the predictors and socio-economic outcomes of young parenthood in Great Britain. *Res Soc Stratif Mobil.* 2007;25(3):205-218. doi:10.1016/j.rssm.2007.08.002
51. Manning WD, Trella D, Lyons H, Toit NC Du. Marriageable Women: A Focus on Participants in a Community Healthy Marriage Program. *Fam Relat.* 2010;59(1):87-102. doi:10.1111/j.1741-



3729.2009.00588.x

52. Sandberg-Thoma SE, Dush CMK. Indicators of Adolescent Depression and Relationship Progression in Emerging Adulthood. *J Marriage Fam.* 2014;76(1):191-206. doi:10.1111/jomf.12081
53. Trillingsgaard T, Sommer D. Associations between older maternal age, use of sanctions, and children's socio-emotional development through 7, 11, and 15 years. *Eur J Dev Psychol.* 2018;15(2):141-155. doi:10.1080/17405629.2016.1266248
54. Dolan C V, Geels L, Vink JM, et al. Testing Causal Effects of Maternal Smoking During Pregnancy on Offspring's Externalizing and Internalizing Behavior. *Behav Genet.* 2016;46(3):378-388. doi:10.1007/s10519-015-9738-2
55. Saha S, Barnett AG, Buka SL, McGrath JJ. Maternal age and paternal age are associated with distinct childhood behavioural outcomes in a general population birth cohort. *Schizophr Res.* 2009;115(2-3):130-135. doi:10.1016/j.schres.2009.09.012

## SUPPLEMENTARY MATERIAL

**Table S1:** Mean and SD for Externalizing and Internalizing Problems of Girls

Rater	Cohort	Externalizing	Internalizing	N-Ext/N-Int
Child	Gen-R	1.77 (1.81)	2.35 (2.15)	2,054/2,058
	RADAR-Y	10.19 (6.75)	0.15 (0.93)	213/120
	TRAILS	7.55 (5.40)	12.06 (7.51)	1,115/1,106
Mother	Gen-R	3.34 (4.29)	4.90 (4.96)	2,305/2,305
	NTR	4.79 (5.33)	4.81 (5.14)	5,626/5,577
	TRAILS	7.26 (6.21)	7.89 (6.25)	1,006/1,002
Father	Gen-R	3.38 (4.20)	4.54 (4.67)	1,655/1,656
	NTR	4.03 (4.84)	3.62 (4.25)	3,764/3,734
Teacher	NTR	2.12 (4.28)	4.29 (4.85)	3,314/3,268
	TRAILS	0.26 (0.59)	0.96 (1.09)	992/993

**Table S2:** Mean and SD for Externalizing and Internalizing Problems of Boys

<b>Rater</b>	<b>Cohort</b>	<b>Externalizing</b>	<b>Internalizing</b>	<b>N-Ext/N-Int</b>
<b>Child</b>	Gen-R	2.12 (2.02)	1.95 (2.00)	1,955/1,959
	RADAR-Y	10.93 (7.44)	-0.20 (0.77)	278/146
	TRAILS	9.85 (6.83)	10.47 (7.23)	1,073/1,065
<b>Mother</b>	Gen-R	4.51 (5.41)	4.81 (5.14)	2,244/2,245
	NTR	6.46 (6.73)	4.55 (4.99)	5,460/5,409
	TRAILS	9.59 (7.62)	7.80 (6.16)	959/953
<b>Father</b>	Gen-R	4.61 (5.48)	4.63 (4.77)	1,604/1,603
	NTR	5.32 (5.87)	3.50 (4.22)	3,656/3,640
<b>Teacher</b>	NTR	4.48 (6.96)	4.52 (5.08)	3,222/3,178
	TRAILS	0.63 (0.88)	1.03 (1.14)	933/931

**Table S3:** Parental Age Predicting Externalizing Problems from Exploratory Results

Rater	Cohort	Age F.		r <sup>2</sup>	Age M.		r <sup>2</sup>
		$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)		$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	
<b>Child</b>	Gen-R	-0.06 (<.007)	.08 (<.001)	.01	-.05 (.02)	.05 (.03)	.01
	RADAR-Y	-.05 (.44)	.14 (.05)	.02	-.08 (.22)	.18 (.01)	.04
	TRAILS	-.01 (.83)	-.01 (.77)	.00	-.03 (.39)	-.03 (.36)	.00
<b>Mother</b>	Gen-R	-.10 (<.001)	.09 (<.001)	.01	-.10 (<.001)	.02 (<.001)	.02
	NTR	-.12 (<.001)	.08 (<.001)	.01	-.11 (<.001)	.06 (<.001)	.02
	TRAILS	.09 (.02)	.08 (.04)	.01	-.13 (<.001)	.06 (.06)	.02
<b>Father</b>	Gen-R	-.10 (<.001)	.08 (.003)	.01	-.08 (.001)	.07 (<.001)	.01
	NTR	-.13 (<.001)	.07 (<.001)	.02	-.12 (<.001)	.06 (<.001)	.02
<b>Teacher</b>	NTR	-.05 (<.001)	.03 (.047)	.00	-.04 (.001)	.04 (.009)	.00
	TRAILS	-.08 (.03)	.06 (.11)	.01	-.11 (<.001)	.04 (.20)	.01

F. = Father.  
M. = Mother

**Table S4:** Age Father and Covariates Predicting Externalizing Problems from Exploratory Results

Rater	Cohort	Age	Age <sup>2</sup>	SES	Gender Child	<i>r</i> <sup>2</sup>
		$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	
Child	Gen-R	-.05 (.03)	.07 (<.001)	-.07 (<.001)	-.08 (<.001)	.02
	RADAR-Y	-.05 (.50)	.13 (.07)	-.06 (.39)	-.06 (.39)	.02
	TRAILS	-.01 (.88)	-.01 (.77)	-.01 (.67)	-.18 (<.001)	.03
Mother	Gen-R	-.09 (<.001)	.07 (.004)	-.08 (<.001)	-.15 (<.001)	.04
	NTR	-.10 (<.001)	.07 (<.001)	-.08 (<.001)	-.13 (<.001)	.05
	TRAILS	-.04 (.27)	.06 (.10)	-.17 (<.001)	-.16 (<.001)	.06
Father	Gen-R	-.10 (<.001)	.06 (.01)	-.06 (.03)	-.15 (<.001)	.04
	NTR	-.11 (<.001)	.06 (<.001)	-.13 (<.001)	-.14 (<.001)	.05
Teacher	NTR	-.04 (.006)	.02 (.125)	-.10 (<.001)	-.17 (<.001)	.04
	TRAILS	-.05 (.20)	.05 (.19)	-.13 (<.001)	-.25 (<.001)	.09

**Table S5:** Age Mother and Covariates Predicting Externalizing Problems from Exploratory Results

Rater	Cohort	Age	Age <sup>2</sup>	SES	Gender Child	r <sup>2</sup>
		β (p-value)	β (p-value)	β (p-value)	β (p-value)	
Child	Gen-R	-.04 (.11)	.04 (.12)	-.06 (.009)	-.08 (<.001)	.02
	RADAR-Y	-.07 (.25)	.17 (.01)	-.05 (.43)	-.05 (.41)	.04
	TRAILS	-.02 (.51)	-.02 (.58)	-.01 (.79)	-.18 (<.001)	.03
Mother	Gen-R	-.08 (<.001)	.06 (.004)	-.06 (.006)	-.14 (<.001)	.04
	NTR	-.09 (<.001)	.06 (<.001)	-.12 (<.001)	-.14 (<.001)	.05
	TRAILS	-.08 (.02)	.06 (.06)	-.15 (<.001)	-.16 (<.001)	.07
Father	Gen-R	-.07 (.009)	.06 (.02)	-.05 (.09)	-.15 (<.001)	.04
	NTR	-.10 (<.001)	.05 (<.001)	-.12 (<.001)	-.14 (<.001)	.05
Teacher	NTR	-.03 (.035)	.03 (.019)	-.10 (<.001)	-.17 (<.001)	.04
	TRAILS	-.07 (.03)	.05 (.11)	-.12 (<.001)	-.25 (<.001)	.09

**Table S6:** Exploratory Results for Parental Age Predicting Internalizing Problems

Rater		Age F.		$r^2$	Age M.		$r^2$
		$\beta$ ( $p$ -value)	$\beta$ ( $p$ -value)		$\beta$ ( $p$ -value)	$\beta$ ( $p$ -value)	
Child	Gen-R	-.03 (.001)	.05 (.020)	.00	-.02 (.32)	.04 (.07)	.00
	RADAR-Y	-.03 (.69)	.03 (.76)	.01	-.04 (.64)	.06 (.41)	.01
	TRAILS	.00 (.98)	-.01 (.78)	.00	-.02 (.55)	.03 (.40)	.00
Mother	Gen-R	-.04 (.12)	.06 (.02)	.00	-.06 (.01)	.05 (.05)	.01
	NTR	-.06 (<.001)	.05 (<.001)	.00	-.06 (<.001)	.03 (.022)	.00
	TRAILS	.01 (.81)	.05 (.17)	.00	-.05 (.12)	.04 (.26)	.00
Father	Gen-R	-.05 (.06)	.06 (.02)	.00	-.03 (.21)	.03 (.28)	.00
	NTR	-.07 (<.001)	.04 (.013)	.01	-.07 (<.001)	.02 (.116)	.01
Teacher	NTR	-.01 (.538)	.02 (.301)	.00	-.01 (.719)	.01 (.299)	.00
	TRAILS	-.02 (.56)	.01 (.89)	.00	-.04 (.21)	.04 (.20)	.00

F. = Father.

M. = Mother.

**Table S7:** Age Father and Covariates Predicting Internalizing Problems from Exploratory Results

Rater	Cohort	Age	Age <sup>2</sup>	SES	Gender Child	<i>r</i> <sup>2</sup>
		$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	
Child	Gen-R	-.02 (.47)	.04 (.08)	-.04 (.07)	.10 (<.001)	.02
	RADAR-Y	-.03 (.75)	.05 (.53)	-.05 (.61)	.27 (.001)	.09
	TRAILS	.01 (.84)	-.01 (.72)	-.02 (.53)	.11 (<.001)	.01
Mother	Gen-R	-.02 (<.001)	.03 (.20)	-.10 (<.001)	.00 (.90)	.01
	NTR	-.05 (<.001)	.04 (.001)	-.06 (<.001)	.02 (.081)	.01
	TRAILS	.03 (.48)	.04 (.24)	-.06 (.06)	.04 (.25)	.01
Father	Gen-R	-.04 (.14)	.04 (.09)	-.05 (.04)	-.02 (.44)	.01
	NTR	-.06 (<.001)	.03 (.034)	-.07 (<.001)	-.01 (.495)	.01
Teacher	NTR	-.00 (.846)	.01 (.386)	-.055 (<.001)	-.03 (.007)	.00
	TRAILS	.02 (.49)	-.02 (.54)	-.16 (<.001)	-.01 (.78)	.03



**Table S8:** Age Mother and Covariates Predicting Internalizing Problems from Exploratory Results

Rater	Cohort	Age	Age <sup>2</sup>	SES	Gender Child	<i>r</i> <sup>2</sup>
		$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	$\beta$ ( <i>p</i> -value)	
Child	Gen-R	-.01 (.64)	.04 (.11)	-.04 (.09)	.11 (<.001)	.02
	RADAR-Y	-.02 (.78)	.09 (.23)	-.04 (.62)	.27 (.001)	.09
	TRAILS	-.02 (.63)	-.03 (.27)	-.01 (.66)	.11 (<.001)	.01
Mother	Gen-R	-.03 (.18)	.03 (.25)	-.09 (<.001)	.00 (.88)	.01
	NTR	-.04 (<.001)	.02 (.049)	-.06 (<.001)	.02 (.085)	.01
	TRAILS	-.04 (.30)	.03 (.34)	-.04 (.19)	.03 (.28)	.01
Father	Gen-R	-.02 (.56)	.02 (.52)	-.06 (.03)	.02 (.47)	.01
	NTR	-.05 (<.001)	.02 (.220)	-.07 (<.001)	-.01 (.489)	.01
Teacher	NTR	.00 (.936)	.01 (.434)	-.05 (<.001)	-.03 (.009)	.00
	TRAILS	.01 (.72)	.03 (.33)	-.15 (<.001)	-.01 (.71)	.03



## Chapter III.B

### **General psychopathology, internalising and externalising in children and functional outcomes in late adolescence**

Hannah Sallis, Eszter Szekely, Alexander Neumann, Alexia Jolicoeur-Martineau, Marinus van IJzendoorn, Manon Hillegers, Celia M.T. Greenwood, Michael J Meaney, Meir Steiner, Henning Tiemeier, Ashley Wazana, Rebecca M Pearson and Jonathan Evans

## ABSTRACT

### Background

Internalising and externalising problems commonly co-occur in childhood. Yet, few developmental models of psychopathology appropriately account for this. We develop a model of childhood psychopathology that separates the unique and shared contribution of individual psychological symptoms into specific internalising, externalising and general psychopathology factors and assess how these general and specific factors predict long-term outcomes concerning criminal behaviour, academic achievement and affective symptoms.

### Methods

Data were drawn from independent birth cohorts (ALSPAC, N=11,612; Generation R, N=7,946; MAVAN, N=408). Child psychopathology was assessed repeatedly using a range of diagnostic and questionnaire-based measures, and multiple informants. First, structural equation models were used to assess the fit of hypothesised models of shared and unique components of psychopathology in all cohorts. Once the model was chosen, linear/logistic regressions were used to investigate whether these factors were associated with important outcomes such as criminal behaviour, academic achievement and wellbeing from late adolescence/early adulthood.

### Results

The model that included specific factors for internalising/externalising and a general psychopathology factor capturing variance shared between symptoms regardless of their classification fitted well for all of the cohorts. As hypothesised, general psychopathology factor scores were predictive of all outcomes of later functioning, while internalising factor scores specifically predicted later internalising outcomes. Externalising factor scores, capturing variance not shared by any other psychological symptoms, were not predictive of later outcomes.

### Conclusions

Early symptoms of psychopathology carry information that is syndrome-specific as well as indicative of general vulnerability and the informant reporting on the child. The “general psychopathology factor” might be more relevant for long-term outcomes than specific symptoms. These findings emphasize the importance of considering the co-occurrence of childhood psychological symptoms when considering long-term impact.

## INTRODUCTION

Psychiatric diagnostic nosology reflects efforts to delineate specific criteria for diagnosing distinct mental disorders across the lifespan. With each revised edition of the diagnostic criteria,<sup>1,2</sup> the total number of disorders as well as the number of diagnoses received by each individual is rising, both for children and adults.<sup>3</sup> Almost half of individuals who meet diagnostic criteria for one disorder also meet diagnostic criteria for another.<sup>4</sup> A similar story is seen within self and parent reported questionnaires for emotional and behavioural problems, where scales are strongly correlated. Therefore, it is important to understand what this comorbidity and common variance represents and its relevance for future important outcomes. Our current research question is whether there is a general factor of child psychopathology and does this predict important outcomes?

While childhood psychopathology is traditionally grouped into internalising and externalising disorders there remains considerable comorbidity between these two categories.<sup>5</sup> In addition, the stability of these categories over time is unclear.<sup>6-8</sup> It is common for underlying internalising disorders to manifest as behavioural problems usually attributed to externalising disorders and vice versa, for example, a child could exhibit features of conduct disorder which result from being anxious. This complexity of the relationship between internalising and externalising symptoms can make it difficult to categorise childhood psychopathology, determine aetiology, investigate outcomes and plan interventions.

Understanding the overlap between internalising and externalising symptoms as well the contribution of multiple informants may improve the characterisation and predictive models of childhood psychopathology. This objective is important for improving childhood problems and preventing later adverse outcomes.<sup>9</sup> Early identification of those at risk is essential for prevention strategies.

Structural equation models (SEM) enable us to consider both general psychopathology and more specific dimensions within the same model.<sup>10-13</sup> In this framework, each symptom can both contribute variance that is shared with other symptoms and which is unique to that symptom. The underlying assumption of bifactor SEM models is that the shared variance amongst items represents a common construct (in our case general psychopathology), as well as unique variance to a smaller cluster of items which represents more specific constructs (for example externalising and internalising). These specific constructs represent the unique variance in these items not accounted for by the overall factor. This approach differs from other techniques such as network analysis, which conceptualise psychopathology as a group of interlinked symptoms without any underlying construct. We test a bifactor model of child psychopathology using data from three independent birth cohorts. Having developed the model, we then test the association between childhood psychopathology and later behavioural, educational

and psychological outcomes in adolescence and early adulthood. Given the comorbidity between internalising and externalising problems and limited evidence of stability of these categories overtime, we hypothesise that the general psychopathology factor will be associated with a range of outcomes. However, internalising symptoms will be associated only with psychological symptoms and externalising with behavioural outcomes.

## METHODS

### Studies and Measures

Data used for these analyses were drawn from the Developmental Research of the Environment, Adversity, Mental health, Biological susceptibility and Gender (DREAM BIG) consortium formed in 2016 to investigate the association between prenatal adversity and later childhood mental health outcomes. DREAM BIG consists of 4 longitudinal population cohorts: the Avon Longitudinal Study of Parents and Children (ALSPAC),<sup>14,15</sup> the Generation Rotterdam (Generation R) Study,<sup>16,17</sup> the Maternal Adversity, Vulnerability and Neurodevelopment (MAVAN) project,<sup>18</sup> and the Growing Up in Singapore Towards healthy Outcomes (GUSTO) study<sup>19</sup>. A full description of each cohort can be found in the relevant cohort profiles and in the supplementary materials. Given that in GUSTO collection of data relevant to the present analysis is still ongoing due to the young age of participants, it was not included in the present study.

Each cohort has collected several measures capturing mental health during early childhood. In the development of a GPF, we focused on those symptoms that quantify common emotional and behavioural problems. Measures included the Development and Wellbeing Assessment, Strengths and Difficulties Questionnaire and the Child Behaviour Checklist. A complete list of measures and full details of each are provided in the supplementary materials.

To maximise the number of participants included in the models and prevent sampling bias, missing information was imputed for participants with available data on at least one psychopathology subscale. Further details on imputation strategies are outlined in the supplementary material. Within ALSPAC, sensitivity analyses were also performed on the subset of participants with complete data on all subscales.

### Modelling psychopathology in childhood

Measures relating to psychopathology from 4 to 8 years of age were collated although each subscale was taken from a single time point within each study. These included self-, parental-, teacher-, and observer-rated measures (Table S1).

Confirmatory factor analysis, a subset of SEM, was used to estimate the general structure of psychopathology, based on previous studies, including one report also based on a subset of data from the Generation R cohort.<sup>13,20</sup> We used a stepwise ap-



proach to construct a model of childhood psychopathology, beginning with a simple unifactor model and building up to a more complex bifactor structure (see Tables 1 and S5 for a complete overview). Model fit was evaluated in each cohort using several model fit indices: root mean square error of approximation (RMSEA), comparative fit index (CFI) and Tucker-Lewis index (TLI). When investigating model fit, RMSEA values of  $<0.05$  and CFI/TLI values of  $>0.9$  are generally used to indicate good fit.

Individual items were first loaded onto a single factor to investigate whether items appeared to be measuring a single construct (unifactor structure). Subsequent models separated the items into specific internalising/externalising factors, defined a priori, to explore whether the items were capturing these two distinct constructs. Most item-scale allocations were known; the few items that did not have a pre-existing allocation, (e.g., the field worker-rated behaviour items in ALSPAC), two researchers independently assigned them based on a priori knowledge (to either the internalising or externalising factor). Although most items loaded strongly onto the factors to which they were initially assigned, some items were moved if modification indices from the initial model indicated that items would be a better fit on the alternative factor (a list of these modifications can be found in the footnote to Table S2).

We also investigated whether additionally accounting for variance common to a specific informant by adding so-called ‘reporter’ factors (i.e., mother, father, teacher, child or field-worker) would further improve model fit (Table 1).

In the final bifactor model, each item loaded onto the GPF, a reporter factor, and its corresponding specific factor (i.e., internalising/externalising) with a few exceptions (with the exception of the SDQ prosocial score, the Social and Communication Disorders Checklist (SCDC), the sleep and ‘other’ sum scores of the Childhood Behaviour Checklist (CBCL), the thought and social problems subscales of the Teacher Report Form (TRF) and the Social Responsiveness Scale (SRS)). The final model solution is displayed in Tables S2-S4. Factors in the final model were defined to be orthogonal.

Analyses were performed using MPlus v.7 in ALSPAC and the lavaan R package in MAVAN and Generation R. Robust maximum likelihood (MLR) estimators were used in the MAVAN and Generation R cohorts, while weighted least square means and variances (WLSMV) were used in ALSPAC. Latent variables were standardized in each of the cohorts.

### **Testing the associations between general and specific factors in the bifactor model and long-term outcomes**

We tested the bifactor model by examining the associations between the general psychopathology, internalising and externalising factors with later outcomes measured in ALSPAC in early adulthood. We compared these to associations with internalising and externalising symptoms in a model without general psychopathology.

Outcomes included: (i) diagnoses of depression and anxiety at 18 years assessed using the Revised Clinical Interview Schedule (CIS-R), (ii) psychological wellbeing assessed

at age 21 using the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS), (iii) criminal activity (defined as any self-reported involvement with the police) at age 21; (iv) alcohol use (defined as any problem drinking) assessed by the Alcohol Use Disorders Identification Test (AUDIT) at age 21, (v) and educational attainment as indicated by receiving a pass grade (C or above) at English or mathematics at GCSE (public exams taken at age 16 in the UK).

Analyses were run using an unadjusted model in addition to a model adjusting for child gender, maternal age at delivery, maternal education and income. These were chosen a priori as measures of adversity that could act as confounders. These were variables that are associated with child emotional/behavioural problems and the later outcomes but not part of the causal pathway.

## RESULTS

A full description of each of the cohorts can be found in the cohort profiles.<sup>14–16,18,19</sup> The final sample size for analysis was 408 in MAVAN, 7,946 in Generation R, and 11,612 in ALSPAC.

### Modelling childhood psychopathology

The unifactor model in each cohort had a poor fit, as did the model with internalising and externalising factors only. Model fit improved with the addition of rater factors and further improved with the inclusion of the GPF. Consistently across all cohorts, the best fitting model was a bifactor solution containing a GPF, the specific internalising/externalising factors, and rater factors. Model fit statistics for all models tested are shown in Table 1.

Initially, the correlation between the internalising and externalising factors was constrained to zero in all models. As a sensitivity analysis, we allowed these factors to correlate. In none of the cohorts, did this substantially improve model fit and the correlation between the internalising-externalising factors was small. Consequently, to ensure consistent and parsimonious models final bifactor models in all cohorts were orthogonal.

The final model structure for ALSPAC, MAVAN and Generation R are displayed in Figure 1 and Tables S2-S4.

### *Sensitivity analysis*

1,129 (9.7%) participants in the ALSPAC cohort had complete data on all items included in the psychopathology model. Analyses were re-run in ALSPAC restricting to this subset of complete cases. A similar pattern was observed, with a bifactor model

**Table 1:** Model fit statistics for final model of childhood psychopathology

	ALSPAC			Generation R			MAVAN		
	RMSEA (90% CI)	CFI	TLI	RMSEA (90% CI)	CFI	TLI	RMSEA (90% CI)	CFI	TLI
<b>Unifactor</b>	0.083 (0.079, 0.087)	0.297	0.274	0.103 (0.102, 0.104)	0.544	0.509	0.084 (0.082, 0.086)	0.460	0.440
<b>Internalising &amp; externalising</b>	0.082 (0.078, 0.086)	0.311	0.289	0.124 (0.123, 0.126)	0.324	0.287	0.082 (0.079, 0.084)	0.544	0.526
<b>Bifactor – int, ext, rater &amp; GPF</b>	0.036 (0.036, 0.036)	0.876	0.863	0.048 (0.047, 0.049)	0.915	0.894	0.055 (0.052, 0.057)	0.787	0.763

containing a GPF, specific internalising/externalising factors, and observer factors found to be the best solution (Table S6).

### Testing the associations between general and specific factors in the bifactor model and long-term outcomes

Results showed that the general psychopathology was associated with a range of different outcomes (Table 2). Specifically, there was evidence of an association between the GPF and: developing a depressive disorder ( $\beta=0.117$ ,  $p=0.001$ ), experiencing decreased psychological wellbeing at age 21 ( $\beta=-0.062$ ,  $p=0.001$ ), failing mathematics ( $\beta=-0.235$ ,  $p<0.001$ ) or English GCSE at age 16 ( $\beta=-0.260$ ,  $p<0.001$ ). Unexpectedly there was an association between GPF and reduced risk of problem drinking ( $\beta=-0.102$ ,  $p<0.001$ ) but no evidence of an association with criminal activity and one with anxiety. In the same bifactor model the specific internalising factor was associated with increased risk for depression ( $\beta=0.085$ ,  $p=0.030$ ) and anxiety ( $\beta=0.184$ ,  $p<0.001$ ), decreased wellbeing ( $\beta=-0.089$ ,  $p<0.001$ ), and failure at mathematics GCSE ( $\beta=-0.054$ ,  $p=0.017$ ). There was little evidence of an association with later problem drinking, criminal behaviour or English GCSE results. There was no evidence of an association between the externalising



**Table 2:** Association between childhood psychopathology and later outcomes adjusted for maternal age at delivery, maternal education, household income and child gender

	Factor	N	INT/EXT model		Bifactor model	
			(no GPF)		(INT, EXT, GPF)	
			Estimate	P-value	Estimate	P-value
Depressive disorder	INT	4260	0.106	0.013	0.085	0.030
	EXT		0.145	<0.001	-0.027	0.497
	GPF		-	-	0.117	0.001
Anxiety	INT	4260	0.204	<0.001	0.184	<0.001
	EXT		0.085	0.063	-0.064	0.147
	GPF		-	-	0.069	0.080
Wellbeing	INT	4205	-0.100	<0.001	-0.089	<0.001
	EXT		-0.079	<0.001	-0.025	0.267
	GPF		-	-	-0.062	0.001
Problem drinking	INT	3654	-0.054	0.065	-0.040	0.158
	EXT		-0.114	<0.001	-0.080	0.010
	GPF		-	-	-0.102	<0.001
Crime	INT	3684	-0.017	0.641	-0.022	0.529
	EXT		0.073	0.035	0.062	0.075
	GPF		-	-	0.050	0.085
Mathematics GCSE – pass grade (C or above)	INT	6081	-0.097	<0.001	-0.054	0.017
	EXT		-0.308	<0.001	0.050	0.055
	GPF		-	-	-0.235	<0.001
English GCSE – pass grade (C or above)	INT	6201	-0.032	0.294	0.015	0.533
	EXT		-0.383	<0.001	0.082	0.001
	GPF		-	-	-0.260	<0.001

factor scores from the bifactor model and adverse outcomes but some evidence of association with a lower risk for later problem drinking ( $\beta=-0.080$ ,  $p=0.010$ ) and better performance at both mathematics ( $\beta=0.050$ ,  $p=0.055$ ), and English GCSE ( $\beta=0.082$ ,  $p=0.001$ ).

In contrast when not including the GPF in the model, externalising factor was associated with increased criminality, depression, anxiety, failure at both mathematics and English GCSE, decreased wellbeing and lower problem drinking (Table 2). The internalising factor showed similar associations with depression, anxiety, wellbeing and reduced

attainment in mathematics. These associations were stronger in the absence of a general psychopathology factor.

Full results for the adjusted models are given in Tables 2 and unadjusted models in S7.

## DISCUSSION

Here we replicated a bifactor model structure of childhood psychopathology in three international birth cohorts in the DREAM BIG consortium. In each cohort, this bifactor model included a specific internalising and externalising factor, as well as a general psychopathology factor representing variance common to all psychological symptoms.

Having replicated this bifactor model structure across three cohorts, we were able to examine the extent to which this factor was associated with long-term follow up data from ALSPAC. As hypothesised the GPF was associated with a range of outcomes. However, the specific internalising factor still predicted depression, anxiety and well-being when accounting for general psychopathology. In contrast the externalising factor which showed some associations in the simpler model was no longer predictive of adverse outcomes once general psychopathology was taken into account.

This suggests that shared variance between externalising and internalising symptoms may be more important for long term outcomes than specific externalising symptoms. However, these results should be replicated in independent cohorts. If this finding does hold, this does not imply that externalising symptoms are not associated with later functioning, rather, that once the shared variance between externalising and internalising is taken into account (i.e., in the form of the GPF), the remaining unique variance does not relate to the examined outcomes of adolescent/adult functioning. This finding is consistent with those of Brikell and colleagues who investigated the association between a general psychopathology factor model and genetic risk scores for attention-deficit hyperactivity disorder.<sup>21</sup> Simply put, the shared variance in the GPF represents children having both externalising and internalising symptoms and the specific factors representing children with 'pure' symptoms. Thus, our results suggest that those at greater risk of later adverse outcomes such as poor school performance are likely to present with both behavioural and emotional symptoms. Identifying these children would enrich our understanding of the developmental pathways which could inform intervention or prevention strategies, such as the development of a universal therapy or repurposing existing therapies in a transdiagnostic approach.<sup>22,23</sup>

Our results also highlighted the importance of accounting for variation common to a specific informant, as this further improved model fit in each cohort. This partially reflects the individual differences inherent in how different informants answer specific items but it also reflects the fact that raters generally complete entire questionnaires. Thus the different rater factors also likely captured questionnaire-specific variance. In

sum, the informant does have a unique contribution to the child's symptom scores, which is important to account for in data analysis. We therefore recommend that developmental researchers collect data from multiple informants, whenever possible.

There are a number of limitations to our analysis that should be considered. First, the measures of psychopathology partially differed across the cohorts and child self-reports were unavailable in ALSPAC for this age group. However, each cohort used a broad range of measures to capture childhood psychopathology and a comparable model solution was found to be the best across all cohorts. Second, there were missing data in each cohort. In order to maximise power and reduce sampling bias we imputed missing data for all participants with available observation on at least one psychopathology subscale. Importantly, consistent results emerged in the sensitivity analysis conducted in ALSPAC with complete cases only. We did not impute outcomes in ALSPAC so were unable to check how outcomes from our prediction models compared with imputed data. Third, different statistical programmes and imputation strategies were used across the cohorts, however our conclusions about which was the best model were consistent despite these differences. Finally, these analyses were based on data from convenient time points in all cohorts thus do not inform us regarding the trajectory of symptoms of internalising and externalising disorders over time. However, we were able to identify a comparable factor structure of early childhood psychopathology across three independent cohorts.

## CONCLUSION

We suggest that models of childhood psychopathology should account for the co-occurrence of emotional and behavioural symptoms, as well as variance specific to these symptoms, and the informant reporting on the child. Our findings further indicate that this co-occurrence of externalising and internalising symptoms may be more informative for the prevention of long-term adverse outcomes than specific symptoms. However, this finding should be replicated in further studies.

## REFERENCES

1. Association AP, others. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub; 2013.
2. World Health Organization. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. 1993. [http://books.google.nl/books?hl=sv&lr=&id=QiWPj\\_23e-i4C&oi=fnd&pg=PR5&dq=ICD-10&ots=EnrbaWv5TI&sig=hE5SfoufoYOayRapKy2rfy9-Ec8](http://books.google.nl/books?hl=sv&lr=&id=QiWPj_23e-i4C&oi=fnd&pg=PR5&dq=ICD-10&ots=EnrbaWv5TI&sig=hE5SfoufoYOayRapKy2rfy9-Ec8). Accessed November 18, 2013.
3. Insel TR. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. *Am J Psychiatry*. 2014;171(4):395-397.
4. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):617-627. doi:10.1001/archpsyc.62.6.617
5. Angold A, Costello EJ, Erkanli A, Angold A, Costello EJ. Comorbidity. *J child Psychol psychiatry*. 1999;40(01):57-87. doi:10.1038/186935d0
6. Murray AL, Eisner M, Ribeaud D. The Development of the General Factor of Psychopathology 'p Factor' Through Childhood and Adolescence. *J Abnorm Child Psychol*. 2016;5. doi:10.1007/s10802-016-0132-1
7. Rutter M, Kim-Cohen J, Maughan B. Continuities and discontinuities in psychopathology between childhood and adult life. *J Child Psychol Psychiatry*. 2006;47(3-4):276-295.
8. Shevlin M, McElroy E, Murphy J. Homotypic and heterotypic psychopathological continuity: a child cohort study. *Soc Psychiatry Psychiatr Epidemiol*. 2017;52(9):1135-1145.
9. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. *The Lancet Psychiatry*. 2016;3(2):171-178.
10. Caspi A, Houts RM, Belsky DW, et al. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci*. 2014;2(2):119-137. doi:10.1177/2167702613497473
11. Laceulle OM, Vollebergh W a. M, Ormel J. The Structure of Psychopathology in Adolescence: Replication of a General Psychopathology Factor in the TRAILS Study. *Clin Psychol Sci*. 2015:1-11. doi:10.1177/2167702614560750
12. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is There a General Factor of Prevalent Psychopathology During Adulthood? *J Abnorm Psychol*. 2012;121(4):971-977. doi:10.1037/a0028355
13. Neumann A, Direk N, Crawford AA, et al. The low single nucleotide polymorphism heritability of plasma and saliva cortisol levels. *Psychoneuroendocrinology*. 2017;85:88-95. doi:10.1016/j.psyneuen.2017.08.011
14. Boyd A, Golding J, Macleod J, et al. Cohort profile: The 'Children of the 90s'-The index offspring of the avon longitudinal study of parents and children. *Int J Epidemiol*. 2013;42(1):111-127. doi:10.1093/ije/dys064
15. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2012;42(1):97-110.
16. Kooijman MN, Kruithof CJ, van Duijn CM, et al. The Generation R Study: design and cohort update 2017. *Eur J Epidemiol*. 2016;31(12):1243-1264. doi:10.1007/s10654-016-0224-9
17. Tiemeier H, Velders FP, Szekeley E, et al. The Generation R Study: A Review of Design, Findings to

- Date, and a Study of the 5-HTTLPR by Environmental Interaction From Fetal Life Onward. *J Am Acad Child Adolesc Psychiatry*. 2012;51(11):1119-1135.e7. doi:10.1016/j.jaac.2012.08.021
18. O'Donnell KA, Gaudreau H, Colalillo S, et al. The maternal adversity, vulnerability and neuro-development project: theory and methodology. *Can J Psychiatry*. 2014;59(9):497-508.
  19. Soh S-E, Chong Y-S, Kwek K, et al. Insights from the growing up in Singapore towards healthy outcomes (GUSTO) cohort study. *Ann Nutr Metab*. 2014;64(3-4):218-225.
  20. Lahey BB, Rathouz PJ, Keenan K, Stepp SD, Loeber R, Hipwell AE. Criterion validity of the general factor of psychopathology in a prospective study of girls. *J Child Psychol Psychiatry Allied Discip*. 2014:1-8. doi:10.1111/jcpp.12300
  21. Brikell I, Larsson H, Lu Y, et al. The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. *bioRxiv*. 2017. doi:http://dx.doi.org/10.1101/193573
  22. Caspi A, Moffitt TE. All for one and one for all: Mental disorders in one dimension. *Am J Psychiatry*. 2018;175(9):831-844.
  23. Krueger RF, Eaton NR. Transdiagnostic factors of mental disorders. *World Psychiatry*. 2015;14(1):27-29. Supplementary info

## SUPPLEMENTARY MATERIALS AND METHODS

### Studies

#### *Avon Longitudinal Study of Parents and Children (ALSPAC)*

ALSPAC is a longitudinal pregnancy cohort which aimed to recruit all pregnant women in the former county of Avon with an expected due date between April 1991 and December 1992. Detailed information has continued to be collected on mothers, partners and children in the cohort, this process has been described in detail elsewhere.<sup>1,2</sup> Out of the 14,541 mothers who entered the study, 11,612 children had data available on at least one psychopathology subscale at age 7 years. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. A fully searchable data dictionary with information on all available measures is available at <http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>.

#### *Generation Rotterdam (Generation R)*

Generation R is a population-based birth cohort with the aim to identify early environmental and genetic determinants of development and health.<sup>3,4</sup> Mothers living in Rotterdam and delivery date between April 2002 and January 2006 were eligible for the study. Out of the 9901 children who entered the study, 7946 had information on at least one psychopathology subscale available at ages 6-8. All analyses are based on this sample. Parents gave informed consent for their children's participation. The Generation R Study is conducted in accordance with the World Medical Association Declaration of Helsinki and study protocols have been approved by the Medical Ethics Committee of the Erasmus Medical Centre, Rotterdam.

#### *Maternal Adversity, Vulnerability and Neurodevelopment study (MAVAN)*

The Maternal Adversity, Vulnerability and Neurodevelopment (MAVAN) study is a Canadian community-based birth cohort. Pregnant women were recruited from obstetric clinics in hospitals from Montreal (Quebec) and Hamilton (Ontario) if they were 18 years of age or older and fluent in either French or English. Greater details about the cohort are provided elsewhere.<sup>5</sup> The DAWBA was designed for use in samples of 5-16 year olds. The DAWBA was rated by parents and teachers in the ALSPAC cohort at age 7 years.

**Social and Communication Disorders Checklist (SCDC).** The Social and Communication Disorders Checklist (SCDC) is a 12-item screening tool for autistic traits/developmental disorders.<sup>7</sup> The SCDC is a parent-reported measure ranging from 0 to 24, with higher scores indicative of more autistic traits. 9 of the items measure traits relating to social interaction and communication skills, with the remaining 3 items measuring behavioural problems and functional impairment. Parents are asked to rate each state-

ment according to behaviour in the previous 6 months as ‘not true’, ‘quite or sometimes true’ or ‘very or often true’, with corresponding scores of 0, 1 and 2. The SCDC was rated by parents in the ALSPAC cohort when children were aged 7.5 years.

**Additional teacher questions.** Additional questions were included within the teacher-rated questionnaires in ALSPAC which assessed the number of troublesome and awkward behaviours, attention and activity, and the burden of these behaviours on the child. These were measured when the children were around 7 years of age.

**Field worker rated observations.** ALSPAC participants attended a clinic at age 7, at which they completed the following 7 sessions: coordination, hearing, allergy, biological samples, measurements and body statistics, vision, and word skills. After each of these sessions, the field worker was asked to rate the child on each of the following attributes: cooperative, shy, fidget, active, attention and responsive/rapport. Each attribute was measured on a scale of 1-3, for example 1 = cooperative, 2 = somewhat cooperative, 3 = uncooperative.

**Social Responsiveness Scale (SRS).** Autistic-like traits were measured using a validated short-form of the SRS.8 The primary caregiver (91% mothers) rated autistic-like traits when children were 6 years ( $M=6.2$ ,  $SD=0.5$ ). The subscales Social Cognition, Social Communication, and Autistic Mannerism were calculated.

**Teachers Rating Form (TRF).** At age 7 years ( $M=6.7$ ,  $SD=1.3$ ) teachers assessed child psychological problems with the TRF 6-18,9 which includes the subscales: Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behaviour, and Aggressive Behaviour.

**Berkeley Puppet Interview (BPI).** Self-reported behaviour problems were measured in Generation R using the BPI, a semi-structured interactive interview<sup>10</sup> conducted at age 6 years ( $M=6.2$ ,  $SD=0.5$ ). During the interview two hand puppets made opposite statements and the child had to choose which statement fit them best. Scoring was performed with video tapes with high intercoder reliability<sup>11</sup> as used to obtain standardized parent reports of common emotional and behavioural problems. For the current study, we included the seven empirically derived narrowband syndrome scales of the CBCL: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems and Aggressive Behaviour. Items were rated on a 3-point scale (not true, somewhat/sometimes true and very/often true). Within MAVAN, the CBCL was completed by the mother at two time points, i.e., at age 4 and 5 years. In Generation R, questionnaires were completed by the primary caregiver (92% mothers) when children were on average 6 years ( $M=6.1$ ,  $SD=0.5$ ).

**Strengths and Difficulties Questionnaire (SDQ).** The SDQ is a reliable and valid brief measure of prosocial behaviour and psychopathology in children.<sup>6</sup> The SDQ asks about 25 attributes of the child, both positive and negative. Each item can be marked as ‘not true’, ‘somewhat true’ or ‘certainly true’. The measure comprises five subscales, each with five items: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems and prosocial behaviour. The MAVAN cohort included maternal ratings



of the SDQ at ages 5 and 6 years and paternal ratings at 5 years. Within Generation R, the same collection of questionnaires that contained the CBCL, were used to assess items from the SDQ prosocial behaviour scale. The SDQ was also assessed in ALSPAC, with parent and teacher ratings collected at around 7 years of age.

**Conners' Parent Rating Scale–Revised: Short Form (CPRS-R).** Within the MAVAN study, ratings of children's externalizing problems were measured using both maternal and paternal ratings on the CPRS-R:S13 at two time points: at five and six years of age. In Generation R, the CPRS-R was completed by the primary caregiver (90% mothers) when children were 8 years ( $M=8.2$ ,  $SD=0.2$ ). The CPRS-R is a well-validated questionnaire for the assessment of ADHD and ODD. Items were rated on a 4-point scale from 0 (not true at all) to 3 (very much true). Three scales of the CPRS-R were used: Inattention/Cognitive Problems, Hyperactivity and Oppositional.

**The Preschool Age Psychiatric Assessment (PAPA).** The PAPA is a semi-structured researcher-administered diagnostic parent interview, feasible and validated for children under age 7.14 Participants reported on the presence or absence (yes/no) of symptoms of seven disorders with the help of standardized drawings. The symptom scales include Major Depressive Disorder, Generalized Anxiety Disorder, Specific Phobias, Social Phobia, ADHD, ODD, and CD. The present study did not include children's ratings for the ADHD subscale, given its low reliability at this young age according to the manual.<sup>16</sup>

### Imputation strategy

For the ALSPAC cohort, imputation was performed using the `ice` command implemented in Stata 14. Variables included in the imputation model included each variable in the P-factor, plus all earlier measures of these variables and auxiliary variables deemed to be related to missingness. Auxiliary variables included maternal and paternal socioeconomic status and level of education, maternal and paternal alcohol use, marital status, perinatal depression, drug use, domestic violence, partner affection and aggression, gestational age, maternal age at delivery, ethnicity, gestation at enrolment, child temperament. 40 imputed datasets were created, and the parameter estimates for each imputation were combined using Rubin's rules as applied by the 'imputation' package in Mplus.

In MAVAN and Generation R, missing data points were estimated using the full information maximum likelihood (FIML) function specified within lavaan. This function uses all available data on the subscales that were included in the model to estimate missing values.



**Table S1:** Summary of measures across cohorts

Rater	ALSPAC		Gen R		MAVAN	
	Measure (Age)	Subscale	Measure (Age)	Subscale	Measure (Age)	Subscale
Parent	<b>DAWBA</b>	Depression	<b>CBCL</b>	Emotionally reactive	<b>CBCL</b>	Emotionally reactive
	(7 years)	General anxiety	(6 years)	Anxious/depressed	(Mother – 4 and 6 years)	Anxious/depressed
		Separation anxiety		Somatic complaints		Somatic complaints
		Social phobia		Withdrawn		Withdrawn
		Specific phobia		Sleep problems		Sleep problems
		ADHD		Attention problems		Attention problems
		Conduct disorder		Aggressive behaviour		Aggressive behaviour
		ODD		Sum score of other items		
	<b>SDQ</b>	Emotional problems			<b>SDQ</b>	Emotional problems
	(7 years)	Peer problems			(Mother – 5 and 6 years; Father – 5 years)	Peer problems
		Conduct problems				Conduct problems
		Hyperactivity				Hyperactivity
		Prosocial				Prosocial
			<b>CPRS-R</b>	ADHD inattentive	<b>CPRS-R</b>	ADHD inattentive
			(8 years)	ADHD hyperactive impulsive	(Mother and father – 5 and 6 years)	ADHD hyperactive impulsive
				ODD		ODD
			<b>SRS</b>	Social cognition	<b>PAPA</b>	Separation anxiety

**Table S1:** (Continued)

Rater	ALSPAC		Gen R		MAVAN	
	Measure (Age)	Subscale	Measure (Age)	Subscale	Measure (Age)	Subscale
			(6 years)	Social communication	(Mother – 6 years)	GAD
				Autistic mannerism		Social phobia
	<b>SCDC</b>	-				Overanxious disorder
	(7.5 years)					Panic disorder
						Depression & dysthymia
						ADHD CD ODD
Teacher	<b>DAWBA</b>	ADHD	<b>TRF</b>	Anxious/depressed		
	(7 years)	Conduct disorder	(7 years)	Withdrawn/depressed		
		ODD		Somatic complaints		
	<b>SDQ</b>	Emotional problems		Social problems		
	(7 years)	Peer problems		Thought problems		
		Conduct problems		Attention problems		
		Hyperactivity		Rule-breaking behaviour		
		Prosocial		Aggressive behaviour		
	<b>Additional questions</b>	Activity symptoms score				



**Table S1:** (Continued)

Rater	ALSPAC		Gen R		MAVAN	
	Measure (Age)	Subscale	Measure (Age)	Subscale	Measure (Age)	Subscale
	(7 years)	Attention symptoms score Burden of attention/activity Awkward behaviours score Troublesome behaviours Burden of troublesome behaviours				
Field worker	<b>Field worker observations</b> (7 years)	Cooperative Fidget Active Attention Responsive				
Child			<b>BPI</b> (6 years)	Depression Separation anxiety Overanxious Oppositional defiant Overt hostility Conduct problems	<b>Dominic</b> (6 years)	Depression Separation anxiety Overanxious Oppositional defiant Conduct disorder Phobias

DAWBA – Development and Well-Being Assessment; ADHD – Attention deficit hyperactivity disorder; ODD – oppositional defiant disorder; 1 CBCL - Child behaviour checklist; SDQ – Strengths and Difficulties Questionnaire; SCDC - Social and Communication Disorders Checklist; CPRS-R - Conners’ parent rating scale – revised: short-form; <sup>3</sup> SRS - Social responsiveness scale; TRF – Teachers rating form; PAPA – Preschool Age Psychiatric Assessment; BPI – Berkeley puppet interview

**Table S2:** Structure of the bifactor model constructed for the ALSPAC cohort

Internalising factor			Externalising factor			General factor (GPF)
Rater	Measure	Scale	Rater	Measure	Scale	
Parent	DAWBA	Depression	Parent	DAWBA	ADHD	All items plus
		General anxiety			Conduct disorder	SDQ parent and
		Separation anxiety			Oppositional defiant disorder	teacher rated
		Social phobia		SDQ	Conduct problems	prosocial
	SDQ	Emotional problems			Hyperactivity	scores, and
					Peer problems	parent rated
Teacher	SDQ	Emotional problems	Teacher	DAWBA	ADHD	SCDC score
					Conduct disorder	
					Oppositional defiant disorder	
				SDQ	Conduct problems	

**Table S2:** (Continued)

Internalising factor			Externalising factor			General factor (GPF)
Rater	Measure	Scale	Rater	Measure	Scale	
						Hyperactivity
						Peer problems
				Additional questions		Activity symptoms score
						Attention symptoms score
						Burden of attention/activity
						Awkward behaviours score
						Troublesome behaviours
						Burden of troublesome behaviours

**Table S2:** (Continued)

Internalising factor			Externalising factor			General factor (GPF)
Rater	Measure	Scale	Rater	Measure	Scale	
			Field-worker		Cooperative	
					Fidget	
					Active	
					Attention	
					Responsive	

In the initial model, items on the SDQ peer problems subscale were split across internalising and externalising factors, the prosocial subscale was included on the externalising factor. Field worker rated 'responsive' items were included on the internalising factors and the 'shyness' items were included in the model. In the final model, peer problems were included as a single subscale on the externalising factor, the prosocial subscales are included on the GPF only and responsiveness has been moved to the externalising factor. Shyness items have been removed from the model as these were not found to load strongly on any of the factors.

**Table S3:** Structure of the bifactor model constructed for the Generation R cohort

Internalising factor			Externalising factor			General factor (GPF)
Rater	Measure	Scale	Rater	Measure	Scale	
Parent	CBCL	Emotionally reactive	Parent	CBCL	Attention problems	All subscales plus CBCL sleep problems, CBCL sum score of other items, TRF social problems, TRF thought problems, and SRS subscales
		Anxious/depressed			Aggressive behaviour	
		Somatic complaints				
		Withdrawn				
				CPRS-R	ADHD inattentive	
					ADHD hyperactive impulsive	
					ODD	
Teacher	TRF	Anxious/depressed	Teacher	TRF	Attention problems	
		Withdrawn/depressed			Rule-breaking behaviour	
		Somatic complaints			Aggressive behaviour	

**Table S3:** (Continued)

Child	Depression	Child	BPI	Oppositional defiant
	Separation anxiety			Overt hostility
				Conduct problems





**Table S4:** Structure of the bifactor model constructed for the MAVAN cohort

Rater	Internalising factor		Rater	Externalising factor		General factor (GPF)
	Measure	Scale		Measure	Scale	
Mother	CBCL	Emotionally reactive	Mother	CBCL	Attention problems	All items plus CBCL mother-rated sleep problems and SDQ mother-rated and father-rated prosocial behaviour
		Anxious/depressed		PAPA	Aggressive behaviour	
		Withdrawn			ADHD	
		Somatic problems			Oppositional defiant disorder	
Mother and father	PAPA	Social phobia	Mother and father	SDQ	Conduct disorder	All items plus CBCL mother-rated sleep problems and SDQ mother-rated and father-rated prosocial behaviour
		Overanxious			Conduct problems	
		Panic	Hyperactivity			
		Depression	Peer problems			
	SDQ	Emotional problems	Conner's PRS-R	Oppositional problems		
		Separation anxiety		Inattention/cognitive problems		
Child	Dominic	Overanxious	Child	Dominic	Hyperactivity	Oppositional defiant disorder
		Simple phobias			Conduct disorder	
		Depression				

**Table S5:** Model fit statistics for final model of childhood psychopathology

	ALSPAC			Generation R			MAVAN		
	RMSEA	CFI	TLI	RMSEA	CFI	TLI	RMSEA	CFI	TLI
	(90% CI)			(90% CI)			(90% CI)		
Unifactor	0.083 (0.079, 0.087)	0.297	0.274	0.103 (0.102, 0.104)	0.544	0.509	0.084 (0.082, 0.086)	0.460	0.440
Internalising & externalising	0.082 (0.078, 0.086)	0.311	0.289	0.124 (0.123, 0.126)	0.324	0.287	0.082 (0.079, 0.084)	0.544	0.526
Internalising & externalising (correlated)*	0.086 (0.080, 0.084)	0.243	0.218	0.105 (0.103, 0.105)	0.352	0.315	0.081 (0.078, 0.083)	0.559	0.541
Internalising, externalising & rater	0.047 (0.043, 0.050)	0.778	0.763	0.060 (0.058, 0.061)	0.857	0.836	0.061 (0.059, 0.064)	0.754	0.733
Internalising, externalising & rater (correlated)**	0.050 (0.046, 0.054)	0.752	0.735	0.060 (0.058, 0.061)	0.857	0.836	0.061 (0.059, 0.064)	0.754	0.733
Bifactor – int, ext & GPF	0.060 (0.056, 0.064)	0.643	0.619	0.090 (0.089, 0.091)	0.674	0.629	0.072 (0.069, 0.074)	0.620	0.592
Bifactor – int, ext, rater & GPF	0.036 (0.036, 0.036)	0.876	0.863	0.048 (0.047, 0.049)	0.915	0.894	0.055 (0.052, 0.057)	0.787	0.763
Correlated bifactor – int, ext, rater & GPF***	0.036 (0.036, 0.036)	0.872	0.859	0.048 (0.047, 0.049)	0.915	0.894	0.055 (0.052, 0.057)	0.787	0.762

\* Correlation between internalising and externalising factors: ALSPAC=0.286,  $p<0.001$ ; Generation R=0.664,  $p<0.001$ ; MAVAN=0.572,  $p<0.001$

\*\* Correlation between internalising and externalising factors: ALSPAC=0.284,  $p<0.001$ ; Generation R=0.108,  $p=0.002$ ; MAVAN=-0.026,  $p=0.814$

\*\*\* Correlation between internalising and externalising factors: ALSPAC=-0.102,  $p<0.001$ ; Generation R=-0.230,  $p<0.001$ ; MAVAN=-0.051,  $p=0.578$

**Table S6:** Model fit statistics restricting to complete cases in the ALSPAC cohort

	ALSPAC		
	RMSEA	CFI	TLI
Unifactor	0.066 (0.064, 0.067)	0.436	0.418
Rater	0.051 (0.049, 0.052)	0.665	0.653
Instrument	0.069 (0.068, 0.070)	0.370	0.349
Internalising & externalising	0.066 (0.065, 0.067)	0.422	0.404
Internalising, externalising & rater	0.040 (0.039, 0.041)	0.800	0.787

**Table S6:** (Continued)

Bifactor – internalising, externalising & P-factor	0.042 (0.041, 0.043)	0.774	0.759
Bifactor – internalising, externalising, rater & P-factor	0.026 (0.025, 0.028)	0.915	0.906
Correlated bifactor – internalising, externalising, rater & P-factor*	0.026 (0.025, 0.028)	0.914	0.905

---

\* Correlation = -0.157,  $p < 0.001$

**Table S7:** Unadjusted association between childhood psychopathology and later outcomes

	Factor	N	No GPF (unadjusted)		Bifactor model (unadjusted)	
			Estimate	P-value	Estimate	P-value
AUDIT problem drinking	INT	3654	-0.065	0.029	-0.053	0.067
	EXT		-0.108	0.0001	-0.068	0.037
	GPF		-	-	-0.093	0.001
Crime binary	INT	3684	-0.042	0.267	-0.056	0.114
	EXT		0.180	<0.001	0.032	0.383
	GPF		-	-	0.143	<0.001
Wellbeing	INT	4205	-0.108	<0.001	-0.101	<0.001
	EXT		-0.041	0.066	-0.047	0.042
	GPF		-	-	-0.028	0.145
Depressive disorder	INT	4260	0.120	0.004	0.108	0.006
	EXT		0.085	0.041	0.007	0.865
	GPF		-	-	0.063	0.076
Anxiety	INT	4260	0.214	<0.001	0.200	<0.001
	EXT		0.036	0.433	-0.039	0.383
	GPF		-	-	0.022	0.572
Maths GCSE – pass grade (C or above)	INT	6081	0.079	0.004	-0.038	0.112
	EXT		0.352	<0.001	0.040	0.148
	GPF		-	-	-0.269	<0.001
English GCSE – pass grade (C or above)	INT	6201	0.004	0.876	0.041	0.095
	EXT		0.447	<0.001	0.090	0.001
	GPF		-	-	-0.350	<0.001

## REFERENCES

1. Boyd A, Golding J, Macleod J, et al. Cohort profile: The 'Children of the 90s'-The index offspring of the avon longitudinal study of parents and children. *Int J Epidemiol.* 2013;42(1):111-127. doi:10.1093/ije/dys064
2. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol.* 2012;42(1):97-110.
3. Kooijman MN, Kruihof CJ, van Duijn CM, et al. The Generation R Study: design and cohort update 2017. *Eur J Epidemiol.* 2016;31(12):1243-1264. doi:10.1007/s10654-016-0224-9
4. Tiemeier H, Velders FP, Szekely E, et al. The Generation R Study: A Review of Design, Findings to Date, and a Study of the 5-HTTLPR by Environmental Interaction From Fetal Life Onward. *J Am Acad Child Adolesc Psychiatry.* 2012;51(11):1119-1135.e7. doi:10.1016/j.jaac.2012.08.021
5. O'Donnell KA, Gaudreau H, Colalillo S, et al. The maternal adversity, vulnerability and neuro-development project: theory and methodology. *Can J Psychiatry.* 2014;59(9):497-508.
6. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The development and well-being assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol psychiatry.* 2000;41(5):645-655.
7. Skuse DH, James RS, Bishop DVM, et al. Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature.* 1997;387(6634):705.
8. Constantino JN. *Social Responsiveness Scale (SRS), Manual.* (Services WP, ed.). Los Angeles; 2002.
9. Achenbach TM, Rescorla LA. *Manual for the ASEBA School-Age Forms and Profiles.* (University of Vermont Youth, and Families RC for C, ed.). Burlington, VT; 2001.
10. Arseneault L, Moffitt TE, Caspi A, et al. Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. *J Child Psychol Psychiatry.* 2003;44(6):832-848. doi:Doi 10.1111/1469-7610.00168
11. Ringoot AP, Jansen PW, Steenweg-de Graaff J, et al. Young children's self-reported emotional, behavioral, and peer problems: the Berkeley Puppet Interview. *Psychol Assess.* 2013;25(4):1273-1285. doi:10.1037/a0033976
12. Achenbach TM, Rescorla LA. *Manual for the ASEBA Preschool Forms and Profiles.* (University of Vermont Youth, and Families RC for C, ed.). Burlington, VT; 2000.
13. Conners CK, Sitarenios G, Parker JDA, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *J Abnorm Child Psychol.* 1998;26(4):257-268. doi:Doi 10.1023/A:1022602400621
14. Zeanah CH, Egger HL, Smyke AT, et al. Institutional rearing and psychiatric disorders in Romanian preschool children. *Am J Psychiatry.* 2009;166(7):777-785.
15. Valla J-P, Bergeron L, Bidaut-Russell M, St-Georges M, Gaudet N. Reliability of the Dominic-R: A Young Child Mental Health Questionnaire Combining Visual and Auditory Stimuli. *J Child Psychol Psychiatry.* 1997;38(6):717-724.
16. Valla J-P, Bergeron L, Smolla N. The Dominic-R: a pictorial interview for 6-to 11-year-old children. *J Am Acad Child Adolesc Psychiatry.* 2000;39(1):85-93.





## Chapter III.C

### **Single Nucleotide Polymorphism Heritability of a General Psycho- pathology Factor in Children**

Alexander Neumann, MSc; Irene Pappa,  
MD, PhD; Benjamin B. Lahey, PhD; Frank C.  
Verhulst, MD, PhD; Carolina Medina-Gomez,  
PhD; Vincent W. Jaddoe, MD, PhD; Marian  
J. Bakermans-Kranenburg, PhD; Terrie E.  
Moffitt, PhD; Marinus H. van IJzendoorn,  
PhD; Henning Tiemeier, MD, PhD



## ABSTRACT

**Objective:** Co-occurrence of mental disorders is commonly observed, but the etiology underlying this observation is poorly understood. Studies in adolescents and adults have identified a general psychopathology factor associated with a high risk for different psychiatric disorders. We defined a multi-informant general psychopathology factor in school-aged children and estimated its SNP heritability. The goal was to test the hypothesis that child behavioral and emotional problems are under the influence of highly pleiotropic common autosomal genetic variants that non-specifically increase the risk for different dimensions of psychopathology.

**Method:** Children from the Generation R cohort were repeatedly assessed between ages 6-8 years. Child behavior problems were reported by parents, teachers and children. Confirmatory factor analysis estimated a general psychopathology factor across informants using various psychiatric problem scales. Validation of the general psychopathology factor was based on IQ and temperamental measures. Genome-wide Complex Trait Analysis (GCTA) was used to estimate the SNP heritability ( $n=2,115$ ).

**Results:** The general psychopathology factor was associated with lower IQ, higher negative affectivity and lower effortful control, but not with surgency. Importantly, the general psychopathology factor showed a significant SNP heritability of 38% ( $SE=0.16$ ,  $p=0.008$ ).

**Conclusions:** Common autosomal SNPs are pleiotropically associated with internalizing, externalizing and other child behavior problems, and underlie a general psychopathology factor in childhood.

## INTRODUCTION

Co-occurrence of mental disorders is commonly observed in clinical and epidemiological studies.<sup>1,2</sup> Family studies have shown familial co-aggregation for many common disorders such as depressive and anxiety disorders, but also ADHD, conduct problems, and psychosis.<sup>3,4</sup> This coincides with the observation that many risk factors for psychopathology, like stressful life events,<sup>5</sup> are not disorder specific. Research suggests the existence of a general psychopathology factor, which is associated with high risk of developing a broad range of both internalizing and externalizing mental disorders and problems<sup>6,7</sup> in preschool children,<sup>8</sup> school-aged children,<sup>9</sup> adolescents,<sup>10–12</sup> and adults.<sup>6</sup> In one study a latent general factor based on repeated assessments of psychiatric symptoms over a 20 year period explained on average 42% of the disorder variance.<sup>13</sup> In another large multi-ethnic adult sample, a general factor was estimated to explain between 29% to 67%, depending on the diagnosis.<sup>14</sup> The general psychopathology factor might be especially prominent in children, since it has been argued that psychiatric disorders are not as differentiated in young age, although the notion has been recently challenged.<sup>15</sup> Most of the aforementioned studies used DSM oriented scales, however, the general psychopathology factor was also replicated in studies using problem scales/items in general population samples.<sup>12,15</sup> Both assessment approaches show concurrent validity with DSM diagnoses (see e.g. Ebesutani et al.<sup>16</sup>), therefore a general psychopathology factor can be estimated with a variety of instruments.

Given the consistent statistical evidence for common etiological pathways between psychopathology domains, the question arises to what extent these pathways are genetic. Twin studies have indicated that a common genetic factor influences a broad range of behavioral and emotional problems in child and adolescent participants,<sup>7,17–19</sup> which can explain the observed co-occurrence to a large degree. However, since twin studies use familial information, it is unclear which specific genetic mechanisms are at play. It is unclear to which extent this heritability can be explained by common or rare variants, single nucleotide polymorphisms (SNP) or structural variants, additive effects or non-additive effects like epistasis. Thus our goal was to test the hypothesis that the general psychopathology factor represents the additive effects of common SNPs and to estimate the variance explained by these (SNP heritability). Various adult psychiatric diagnoses show substantial genetic correlations (up to 0.68) based on common autosomal variants,<sup>20</sup> however, the SNP heritability of a general psychopathology factor is unknown. To estimate the SNP heritability, we used genomic restricted maximum likelihood estimation (GREML), which estimates heritability by quantifying the extent to which individuals who share more SNP alleles also are more similar phenotypically.<sup>21</sup>

To test the hypothesis we first defined a general psychopathology factors using multi-informant data, i.e. parent, teacher and child self-report measures, in a large population-based cohort: the Generation R study. Psychiatric research in young chil-

dren that relies on a single informant, most often a parent, may bias results and inflate correlations between behavior subscales due to common method variance.<sup>22</sup> For the calculation of the general psychopathology factor score, we extended a validated model,<sup>13</sup> which has been replicated in adolescents,<sup>12</sup> by including autistic-like behaviors and accounting for the multiple informant context. To test validity of our general factor we examined its association with intelligence and temperament. Intelligence is an important criterion variable due to several reasons: it is measured independent of psychopathology and is an indicator of neurodevelopment. Furthermore, low IQ has previously been associated with general factors of psychopathology.<sup>9,13</sup> Higher levels of neuroticism have also been observed with a general risk for psychopathology.<sup>8,13,19</sup>

## METHOD

### Participants

This study features participants from a population-based birth cohort, the Generation R Study,<sup>23</sup> designed to identify early environmental and genetic predictors of development and health from fetal life onwards. All participating children were born between April 2002 and January 2006. 6,624 children were all repeatedly assessed during the ages 6-8 of which 1,954 had complete data on all problem subscales of all instruments and were used in a confirmatory factor analysis (CFA) of the general psychopathology model. We established the associations between the subscales and the factor in the set with complete data and then applied this information to estimate factor scores in the incomplete dataset using multiple imputation (see supplementary methods 1). IQ information was available in 1,826 children with complete problem subscales and temperament information was available in 1,933 children.

The sample used for the non-genetic analyses was multi-ethnic: the majority had a Dutch, Surinamese, Turkish or Moroccan ethnic background. However, for GREML analyses only children with genetic data and of European ancestry were eligible (n=2,353). 238 participants were removed from analysis due to excessive relatedness, resulting in a final GREML sample size of 2,115. All parents gave informed consent for their children's participation and teacher report. Study protocols were approved by the local ethics committee.

### Instruments

The general psychopathology factor was estimated using parent, teacher and self-reported measures, designed to assess common behavioral and emotional problems in the general population. We took a conservative approach to missing data. All children included in any analysis with incomplete data had at least 50% of observations

complete and had been assessed with multiple instruments. See Figure S1 for correlations between all subscales.

### **Child behavior problems assessed by parents**

#### *Child Behavior Checklist (CBCL)*

The CBCL 1 ½ -5 years was used to assess child behavioral problems in pre-school-aged children.<sup>24</sup> Questionnaires were completed by the primary caregiver (92% mothers) when children were on average 6 years ( $M=6.1$ ,  $SD=0.5$ ). The CBCL includes seven subscales: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, Aggressive Behavior; and a sumscore of other items.

#### *Social Responsiveness Scale (SRS)*

The SRS is a parent-report questionnaire assessing autistic-like traits in children.<sup>25</sup> A validated short-form was completed by the primary caregiver (91% mothers) when children were 6 years ( $M=6.2$ ,  $SD=0.5$ ). Three subscales were used: Social Cognition, Social Communication, and Autistic Mannerism.

#### *Conners' Parent Rating Scale-Revised (CPRS-R)*

The CPRS-R is a well-validated questionnaire, assessing attention deficit/hyperactivity disorder (ADHD) and Opposition Defiant Disorder (ODD).<sup>26</sup> It was completed by the primary caregiver (90% mothers) when children were 8 years ( $M=8.2$ ,  $SD=0.2$ ). Three scales of the CPRS-R were used: ADHD Inattentive, ADHD Hyperactive-Impulsive, and ODD.

### **Child behavior problems assessed by teachers**

#### *Teacher's Rating Form (TRF)*

Teachers were approached independently of parents at 7 years ( $M=6.8$ ,  $SD=1.3$ ), and completed the TRF 6-18 years.<sup>27</sup> The following subscales were used: Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior.

### **Child behavioral problems assessed by Children**

#### *Berkeley Puppet Interview (BPI)*

The BPI is a semi-structured interactive interview<sup>28</sup> conducted in our research center with the help of two identical dog hand puppets at 6 years ( $M=6.2$ ,  $SD=0.5$ ). The two puppets made opposite statements and the child was asked to indicate which statement described him/her best. The interview was videotaped and scored with high in-

tercoder reliability.<sup>29</sup> Six subscales were used: Depression, Separation Anxiety, Overanxious, Oppositional Defiant, Overt Hostility, and Conduct Problems.

### **Validation Measures**

Non-verbal cognitive abilities were assessed with the Snijder-Oomen nonverbal intelligence test.<sup>30</sup> This is a well-validated test administered at age 6. The mean IQ score was 103.9 (SD=14.2).

We assessed temperamental dimensions (negative affectivity, surgency/extraversion, and effortful Control) at age 6 with the Child Behavior Questionnaire (Very-Short-Form), a parent-rated questionnaire.<sup>31</sup>

### **Genotyping and Quality Control**

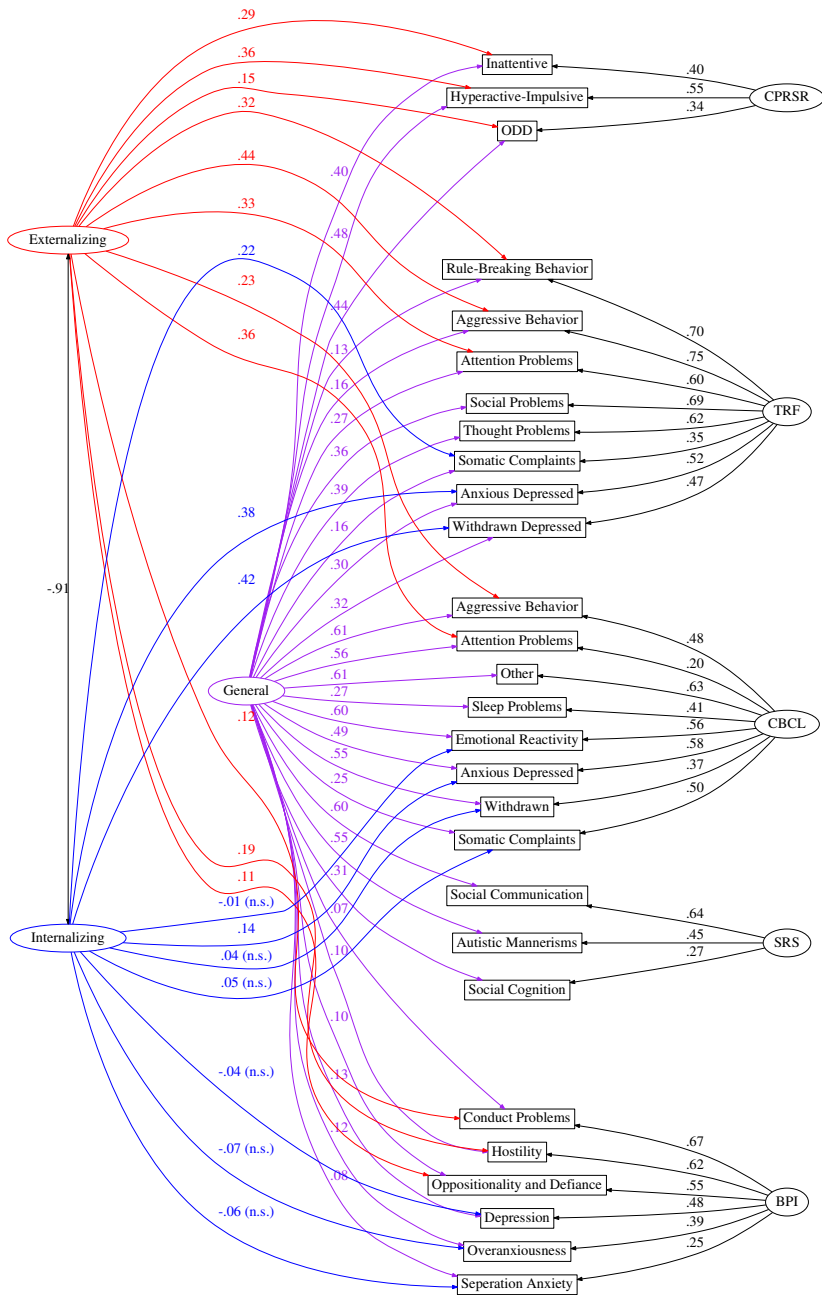
DNA was extracted from whole blood cells from cord blood and genotyped on Illumina 610K/660W platforms. Quality checks for each SNP included sample ( $\geq 97.5\%$ ) and SNP call rates ( $\geq 90\%$ ), minor allele frequency  $\geq 1\%$  and deviation from the Hardy-Weinberg equilibrium ( $p < 10^{-7}$ ). Samples were checked for excess heterozygosity, gender accuracy, relatedness, and missing data. After quality control 504,617 autosomal SNPs remained. Genetic ancestry was investigated using multidimensional scaling. Participants were classified as non-northwestern European ancestry when they exceeded 4 SDs difference with the mean European reference level (HapMap CEU) on any of the first four principal components. 2830 were classified as European ancestry and 2901 as non-northwestern European. Principal components of ancestry used as covariates in this study were based on the European sample. See Medina-Gomez et al. for further details on the genetic data in Generation R.<sup>32</sup>

### **Statistical Analyses**

We performed the statistical analyses in four steps: 1) The general psychopathology factor was estimated with a CFA 2) The general psychopathology factor was associated with IQ and temperament to confirm validity 3) Factor scores were extracted and imputed 4) The SNP heritability of the general factor score was estimated.

#### *General Psychopathology Factor Model*

The continuous subscales of specific child emotional and behavioral problems constituted the manifest (observed) variables. A latent variable loading on all subscales from all instruments was specified to represent the general psychopathology factor. Two other latent factors loaded on internalizing and externalizing problems. These were allowed to correlate, however, the correlations to the general psychopathology factor were constrained to 0. This bifactor model is adapted from the best fitting models found in previous studies.<sup>7,12,13</sup> Consequently, the internalizing and externalizing factors represent effects specific to internalizing and externalizing problems, respectively, that cannot be explained by a general vulnerability to psychopathology. The general psycho-



**Figure 1:** Path diagram of the general psychopathology factor model with standardized loadings. Sex and age paths are not displayed

pathology factor in turn represents a general vulnerability, independent of these more specific effects. Sex and age were included as covariates.

To account for within-instrument bias, latent variables specific to the subscales of each instrument were introduced (Figure 1) that were not allowed to correlate with other latent variables. The assumption is that similarities between subscales measured by the same instrument, which are not shared with subscales of other instruments, capture bias due to common method variance. It is therefore important, that these instrument specific latent variables are not correlated with each other. If they were, they would also capture variances, which are shared with other instruments. These cross-instrument effects, however, should contribute to the general factor of psychopathology.

### *Confirmatory Factor Analysis*

The CFA models were fit in a subsample with complete data on all subscales (n=1,954) using standardized latent variables. Because of violations of assumption of multivariate normality, the maximum likelihood robust estimator was used. The general psychopathology model was formally compared to a simpler model without the general psychopathology factor, but otherwise identical model. Model fit was judged by: comparative fit index (CFI), Tucker-Lewis index (TLI), root-mean-square error of approximation (RMSEA), standardized root mean square residual (SRMR), and Bayesian information criterion (BIC). Llavaan 0.5-20 in R 3.2.3 was used for the CFA.<sup>33,34</sup> Factor scores were extracted from the CFA model, imputed and used for further analyses (see supplementary methods 1).

### *Validation*

To validate the general psychopathology factor, we extended the general psychopathology model to two structural equation models. In one model IQ predicted the general and specific internalizing and externalizing factors and in another model the temperamental dimensions were the predictors.

### *SNP Heritability*

Next, we estimated the variance explained by additive effects of autosomal SNPs using GREML as implemented in GCTA 1.24.7.<sup>21,35</sup> A potential source of bias is the ethnic admixture in the heterogeneous Generation R sample because participants from different ethnicities, and thus distant genetic relatedness, might have concordant or discordant phenotypes due to environmental correlates of ethnicity and not genetic makeup. To minimize population stratification, we restricted analyses to children with European ancestry (2830 of 5731, of which 2353 had information on the phenotype). Moreover, we included four principal components of ancestry, that were estimated within the European sample,<sup>32</sup> as covariates (see supplementary methods 2 for measurement invariance tests between European and non-northwestern European ancestries). To reduce confounding due to shared environment, the conventional GRM cutoff 0.025 was used

to exclude close relatives (second-degree cousins and closer) within the European sample ( $n=238$ ).<sup>21</sup> The final GREML sample with complete genetic information consisted of 2,115 children of European ancestry.

First a genetic relatedness matrix (GRM) between unrelated participants was calculated based on autosomal SNPs. Second, restricted maximum likelihood was used to estimate the phenotype (general psychopathology factor scores) variance explained by the random effect of GRM. SNP heritabilities of the specific internalizing and externalizing factors from the general psychopathology factor model are not presented. Lower item count and loadings of these factors make factor score estimation problematic due to factor indeterminacy, thus SNP heritabilities of such factor scores could be misleading. A likelihood ratio test compared whether the inclusion of the GRM significantly improved model fit. The factor score was transformed using  $\ln(\text{score} - \text{lowest score} + 1)$  to normalize the total genetic effects and residuals distribution. All GREML results were similar for transformed and untransformed factor scores. The total genetic effects were similar for observations with and without imputation (see supplementary methods 1).

## RESULTS

### General Psychopathology Factor Model

All psychopathology subscales of all instruments/raters loaded significantly on the general psychopathology factor independent of the more specific internalizing/externalizing factors and the instrument-specific covariances (Table 1). Furthermore, although BIC penalizes models for greater complexity, the BIC was 1,358 points lower for the model including the general factor as opposed to the same model with the general psychopathology factor omitted, providing strong evidence for a difference in model fit (See Table 2).<sup>36</sup>

The average loading of the individual psychopathology scales on the general factor was 0.34. Parent-rated subscales showed the highest average loadings ( $M=0.48$ ), teacher ratings showed moderate average loadings ( $M=0.26$ ), and child-report subscales showed modest, but still significant average loadings ( $M=0.10$ ). In the hierarchical model with the general factor, the internalizing and externalizing scales correlated strongly and negatively with each other ( $r=-0.91$ ,  $SE=0.18$ ,  $p<0.001$ ). All externalizing subscales loaded significantly on the externalizing factor. Similarly, all teacher-reported internalizing subscales and the parent-reported “Anxious/Depressed” subscale loaded on the internalizing factor. No other internalizing subscale loaded on the internalizing factor in the model including the general psychopathology factor. The instrument-specific latent



**Table 1:** Standardized factor loadings of the general psychopathology factor model

Items	General			
	Loading	SE		SE
<i>Parent Report:</i>				
Emotional Reactivity (CBCL)	0.60***	0.06		-0.01 0.03
Anxious Depressed (CBCL)	0.49***	0.06		0.14*** 0.03
Somatic Complaints (CBCL)	0.25***	0.04		0.05 0.03
Withdrawn (CBCL)	0.55***	0.07		0.04 0.03
Attention Problems (CBCL)	0.56***	0.07	0.36***	0.07
Aggressive Behavior (CBCL)	0.61***	0.09	0.23***	0.05
Sleep Problems (CBCL)	0.27***	0.05		
Other (CBCL)	0.61***	0.05		
Social Cognition (SRS)	0.31***	0.04		
Social Communication (SRS)	0.60***	0.08		
Social Autistic Mannerisms (SRS)	0.55***	0.08		
ODD (CPRS-R)	0.44***	0.07	0.15*	0.07
Inattentive (CPRS-R)	0.40***	0.04	0.29***	0.07
Hyperactive-Impulsive (CPRS-R)	0.48***	0.07	0.36***	0.08

**Table 1:** (Continued)

<i>Teacher Report:</i>					
Anxious Depressed (TRF)	0.30***	0.06		0.38***	0.05
Withdrawn Depressed (TRF)	0.32***	0.07		0.42***	0.06
Somatic Complaints (TRF)	0.16**	0.05		0.22***	0.05
Social Problems (TRF)	0.36***	0.05			
Thought Problems (TRF)	0.39***	0.08			
Attention Problems (TRF)	0.27***	0.05	0.33***		0.04
Rule-Breaking Behavior (TRF)	0.13**	0.04	0.32***		0.06
Aggressive Behavior (TRF)	0.16***	0.05	0.44***		0.05
<i>Child Self Report:</i>					
Depression (BPI)	0.13***	0.03		-0.04	0.05
Seperation Anxiety (BPI)	0.08**	0.03		-0.06	0.04
Overanxiousness (BPI)	0.12***	0.03		-0.07	0.05
Oppositionality and Defiance (BPI)	0.10**	0.03	0.11**		0.04
Hostility (BPI)	0.10**	0.03	0.19***		0.04

**Table 1:** (Continued)

Conduct Problems (BPI)	0.07*	0.03	0.12**	0.04
<i>Specific Internalizing/Externalizing Correlations:</i>				
			r	SE
Specific Int/Ext			-0.91***	0.18
* $p < 0.05$ , ** $p < 0.01$ , *** $p < 0.001$				
$n = 1954$				

factors showed moderate loadings ranging from 0.20 to 0.75 (Figure 1), indicating the magnitude of covariance unique to the measurement context.

### *IQ Associations*

The general psychopathology factor was negatively associated with non-verbal IQ, the standardized path coefficient being  $-0.14$  ( $SE = 0.04$ ,  $p = 0.001$ ). However, it was not statistically significantly associated with the specific internalizing ( $\beta = -0.05$ ,  $SE = 0.05$ ,  $p = 0.371$ ) and externalizing factors ( $\beta = -0.09$ ,  $SE = 0.05$ ,  $p = 0.060$ ).

### *Temperament Associations*

Parent-reported temperamental negative affectivity, was associated with the general psychopathology factor ( $\beta = 0.52$ ,  $SE = 0.07$ ,  $p < 0.001$ ), but not with the specific internalizing ( $\beta = -0.12$ ,  $SE = 0.09$ ,  $p = 0.197$ ) or specific externalizing factors ( $\beta = 0.05$ ,  $SE = 0.08$ ,  $p = 0.537$ ). In contrast, parent-reported surgency showed no association with the general psychopathology factor ( $\beta = -0.09$ ,  $SE = 0.07$ ,  $p = 0.169$ ), but a negative association with the specific internalizing ( $\beta = -0.29$ ,  $SE = 0.10$ ,  $p = 0.004$ ) and a positive association with the specific externalizing factor ( $\beta = 0.62$ ,  $SE = 0.04$ ,  $p < 0.001$ ). Effortful control was negatively associated with the general psychopathology factor ( $\beta = -0.20$ ,  $SE = 0.04$ ,  $p < 0.001$ ), pos-

**Table 2:** Model fit indices of the general psychopathology factor (GPF) and the same model without the general psychopathology factor (No GPF)

Model	CFI	TLI	RMSEA <sup>1</sup>	SRMR	BIC
GPF	0.890	0.854	0.048 [0.046, 0.050]	0.036	283891
No GPF	0.831	0.790	0.057 [0.056, 0.059]	0.104	285249

<sup>1</sup>A 90% confidence interval is given for RMSEA  
n=1954

itively associated with the specific internalizing ( $\beta=0.11$ ,  $SE=0.04$ ,  $p=0.017$ ), and negatively associated with the specific externalizing factor ( $\beta=-0.19$ ,  $SE=0.03$ ,  $p<0.001$ ).

### *SNP Heritability*

We observed significant SNP heritability of the general psychopathology factor (SNP  $h^2 = 38\%$ ,  $SE=0.16$ ,  $p=0.008$ , genetic variance=0.028, residual variance=0.047, phenotypic variance=0.076).

## DISCUSSION

A general psychopathology factor, underlying several diverse dimensions of externalizing, internalizing, and autistic-like behaviors, was observed in early school-age children. No single psychopathology subscale was disproportionately associated with this general factor. Importantly, GREML demonstrated substantial SNP heritability of the general psychopathology factor, suggesting that the shared effects responsible for co-occurrence are partly due to common SNPs. The three-informant design is an important strength of the current study, since it allowed an estimation of the general psychopathology factor with lower risk of informant and context bias. This is a valuable extension of previous studies, which reported on the effects of the general psychopathology factor for one informant only or separately for parental, teacher or self-report. The multi-informant approach, however, reduces the chance that a general factor partly reflects reporting tendencies. Another strength is the simultaneous inclusion of autistic-like behavior together with internalizing and externalizing problems. While a previous study in children reported that a general factor explained various neurodevelopmental symptoms, internalizing behaviors were not included.<sup>18</sup> In the present study,

not only externalizing, but all internalizing problems and three autism-like behavior subscales loaded on the general psychopathology factor.

All child-rated scales loaded significantly on the general psychopathology factor and contributed meaningfully to the multi-informant internalizing and externalizing factors. This supports the view that children can report their problems at this young age. The lower loadings of the child-rated and teacher-rated scales compared to parent-rated scales in the general psychopathology factor model, however, might indicate limitations of these informants. Child developmental level, including short attention span, and difficulties to report on complex constructs, are inherent challenges to obtaining valid self-reports at this young age; whereas teachers may know children less well. Alternatively, the lower loadings in the general psychopathology factor model might reflect our conservative within-instrument covariance correction: instrument-specific factors overcorrect for “true” covariance structures detected using a particular instrument or informant only. For example, high teacher-specific loadings may represent the teachers’ specific insights from task-oriented settings and their ability to compare with other children. Parents have more limited insights in the school setting, but can judge children by their behavior outside of school and overall typically spend much more time with them than other informants. However, we took a conservative approach and argue that the likelihood of a true covariance pattern emerging in only one of many instruments is lower than the likelihood of error and thus deliberately introduced the latent factors.

The criterion validity of the general factor was supported by associations with negative affectivity, effortful control and IQ. The temperamental subscales negative affectivity and effortful control represent the disposition to show distress and the ability to self-control. Both traits have been hypothesized to relate to a broad range of disorders, perhaps due to eliciting maladaptive responses from family and peers or shared genetic origins,<sup>37</sup> and thus a general psychopathology factor was expected to be related. Indeed, previous studies also found evidence for robust associations of the general psychopathology factor with negative affectivity and poor effortful control/conscientiousness.<sup>8,13,19</sup> Temperamental surgency, which describes the tendency to experience positive emotions, was not related to the general factor. Particularly strong support for the criterion validity of our general factor comes from the correlation with IQ scores. If the general psychopathology factor was solely a spurious product of common method variance, it would not be correlated with IQ scores obtained in formal testing at our research center, independently of any questionnaire on psychopathology.

Perhaps most importantly, we observed that additive effects of common autosomal SNP variants underlie the general psychopathology factor. This heritable component was found to explain 38% (95% CI [6%-69%]) of the general psychopathology factor variance in the European ancestry sample. This observed SNP heritability is very unlikely to occur under the assumption that the population heritable component is 0, and thus, the finding strongly supports the hypothesis that the general factor of psychopathology reflects shared genetic influences.<sup>6,7</sup> However, the variance explained by common SNPs



is hard to quantify exactly due to the wide confidence intervals, when estimating SNP heritabilities in samples with a few thousand participants.<sup>38</sup> Substantial residual variance remained, which can reflect environmental effects, non-additive effects, polygenic rare variants,<sup>12,39</sup> or structural variants, such as copy number variations, but which also includes the error term.

We previously reported SNP heritabilities for single parent-rated and teacher-rated instruments separately in participants from the Generation R and the Netherlands Twin Register cohorts.<sup>40</sup> These estimates ranged from 12% (Parent-rated (CBCL) Internalizing and Externalizing scale) to 71% (Teacher-rated (TRF) attention problems). The present sample would have been underpowered to detect genetic effects, if the SNP heritability of the general factor were in the lower range of previous estimates. However, the general psychopathology factor is based on the combined analyses of the previously reported and additional instruments. The combination of instruments and the expected increase in accuracy due to multiple informants formed the background of our hypothesis that the SNP heritability of the general factor would be in the middle or higher regions of previous estimates. This hypothesis was confirmed.

The results of the specific internalizing and externalizing factors in the general psychopathology factor model are intriguing. These factors are not the traditionally observed internalizing and externalizing domain scores as measured by their broadband scales. The factors represent specific influences on problematic behavior beyond those explained by the general vulnerability. While previous studies found negative correlations between the specific factors of approximately  $r=-0.47$  in adults and  $r=-0.44$  in children,<sup>12,13</sup> the correlation in the current study was stronger ( $r=-0.91$ ). This suggests the presence of a single internalizing/externalizing dimension in young children, once we account for the general psychopathology factor. Fitting a model with such a dimension is nearly equivalent to the original fit with two highly correlated factors (see Table S1 and S2). The model implies that children scoring high on the externalizing end have more externalizing problems, less internalizing problems and more surgency (see Table S3) compared to children scoring in the center of the dimension (assuming equal scores on general psychopathology). The reverse patterns holds for children scoring high on the internalizing end; they have more internalizing and less externalizing problems and show less surgency. In contrast, this factor is not related to negative affectivity as is the general psychopathology factor. Such an alternative model needs further investigation and replication.

In conclusion, the results suggest that common autosomal SNPs are pleiotropically associated with internalizing, externalizing, autism-like and other problematic behaviors in children. This may suggest that a substantial portion of the genetic variants underlying psychopathology could be missed in GWAS focusing on single disorders. Future GWAS should therefore incorporate the simultaneous analysis of diverse problems. Ac-

counting for the interrelated nature of psychiatric disorders may help to unravel part of the complex genetic architecture of child psychopathology.

## REFERENCES

1. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):617-627.
2. Angold A, Costello EJ, Erkanli A, Angold A, Costello EJ. Comorbidity. *J child Psychol psychiatry*. 1999;40(01):57-87.
3. Middeldorp CM, Cath DC, Van Dyck R, Boomsma DI. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychol Med*. 2005;35(5):611-624.
4. DeVlyder JE, Oh HY. A Systematic Review of the Familial Co-Aggregation of Schizophrenia With Non-Psychotic Disorders.
5. Kim KJ, Conger RD, Elder GH, Lorenz FO. Reciprocal influences between stressful life events and adolescent internalizing and externalizing problems. *Child Dev*. 2003;74(1):127-143.
6. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is There a General Factor of Prevalent Psychopathology During Adulthood? *J Abnorm Psychol*. 2012;121(4):971-977.
7. Lahey BB, Van Hulle CA, Singh AL, Waldman ID, Rathouz PJ. Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Arch Gen Psychiatry*. 2011;68(2):181-189.
8. Olino TM, Dougherty LR, Bufferd SJ, Carlson G a., Klein DN. Testing Models of Psychopathology in Preschool-aged Children Using a Structured Interview-based Assessment. *J Abnorm Child Psychol*. 2014:1201-1211.
9. Lahey BB, Rathouz PJ, Keenan K, Stepp SD, Loeber R, Hipwell AE. Criterion validity of the general factor of psychopathology in a prospective study of girls. *J Child Psychol Psychiatry Allied Discip*. 2014:1-8.
10. Patalay P, Fonagy P, Deighton J, Belsky J, Vostanis P, Wolpert M. A general psychopathology factor in early adolescence. *Br J Psychiatry*. 2015;207(1):15-22.

11. Stochl J, Khandaker GM, Lewis G, et al. Mood, anxiety and psychotic phenomena measure a common psychopathological factor. *Psychol Med.* 2014;45:1-11.
12. Laceulle OM, Vollebergh W a. M, Ormel J. The Structure of Psychopathology in Adolescence: Replication of a General Psychopathology Factor in the TRAILS Study. *Clin Psychol Sci.* 2015:1-11.
13. Caspi A, Houts RM, Belsky DW, et al. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci.* 2014;2(2):119-137.
14. Kim H, Eaton NR. The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *J Abnorm Psychol.* 2015;124(4):1064-1078.
15. Murray AL, Eisner M, Ribeaud D. The Development of the General Factor of Psychopathology “p Factor” Through Childhood and Adolescence. *J Abnorm Child Psychol.* 2016;5.
16. Ebesutani C, Bernstein A, Nakamura BJ, Chorpita BF, Higa-Mcmillan CK, Weisz JR. Concurrent validity of the child behavior checklist DSM-Oriented Scales: Correspondence with DSM diagnoses and comparison to syndrome scales. *J Psychopathol Behav Assess.* 2010;32(3):373-384.
17. Spatola CAM, Fagnani C, Pesenti-Gritti P, Ogliari A, Stazi M-A, Battaglia M. A general population twin study of the CBCL/6-18 DSM-oriented scales. *J Am Acad Child Adolesc Psychiatry.* 2007;46(5):619-627.
18. Pettersson E, Anckarsäter H, Gillberg C, Lichtenstein P. Different neurodevelopmental symptoms have a common genetic etiology. *J Child Psychol Psychiatry.* 2013;54(12):1356-1365.
19. Tackett JL, Lahey BB, van Hulle C, Waldman I, Krueger RF, Rathouz PJ. Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *J Abnorm Psychol.* 2013;122(4):1142-1153.
20. Lee SH, Ripke S, Neale BM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet.* 2013;45(9):984-994.
21. Yang J, Lee S, Goddard M, Visscher P. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet.* 2011;88(1):76-82.
22. Ringoot AP, Tiemeier H, Jaddoe VWV, et al. Parental depression and child well-being: young children’s self-reports helped addressing biases in parent reports. *J Clin Epidemiol.* 2015.
23. Jaddoe VW V, van Duijn CM, Franco OH, et al. The Generation R Study: design and cohort update 2012. *Eur J Epidemiol.* 2012;27(9):739-756.
24. Achenbach TM, Ruffle TM. The Child Behavior Checklist and Related Forms for Assessing Behavioral/Emotional Problems and Competencies. *Pediatr Rev.* 2000;21(8):265-271.
25. Constantino JN. Social Responsiveness Scale (SRS), Manual. (Services WP, ed.). Los Angeles; 2002.
26. Conners CK, Sitarenios G, Parker JDA, Epstein JN. The revised Conners’ Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. *J Abnorm Child Psychol.* 1998;26(4):257-268.
27. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms and Profiles. . (University of Vermont Youth, and Families RC for C, ed.). Burlington, VT; 2001.
28. Arseneault L, Moffitt TE, Caspi A, et al. Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins’ self-reports. *J Child Psychol Psychiatry.* 2003;44(6):832-848.
29. Ringoot AP, Jansen PW, Steenweg-de Graaff J, et al. Young children’s self-reported emo-



- tional, behavioral, and peer problems: the Berkeley Puppet Interview. *Psychol Assess.* 2013;25(4):1273-1285.
30. Tellegen P, Laros J. The construction and validation of a nonverbal test of intelligence: the revision of the Snijders-Oomen tests. *Eur J Psychol Assess.* 1993;9(2):147-157.
  31. Putnam SP, Rothbart MK. Development of short and very short forms of the Children's Behavior Questionnaire. *J Pers Assess.* 2006;87(1):102-112.
  32. Medina-Gomez C, Felix JF, Estrada K, et al. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *Eur J Epidemiol.* March 2015.
  33. Rosseel Y. {lavaan}: An {R} Package for Structural Equation Modeling. *J Stat Softw.* 2012;48(2):1-36. <http://www.jstatsoft.org/v48/i02/>.
  34. R Core Team. R: A Language and Environment for Statistical Computing. 2015. <http://www.r-project.org>.
  35. Yang J, Lee SH, Goddard ME, Visscher PM. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol.* 2013;1019:215-236.
  36. Raftery A. Bayesian model selection in social research. *Sociol Methodol.* 1995;25(111-163).
  37. Nigg JT. Temperament and developmental psychopathology. *J Child Psychol Psychiatry Allied Discip.* 2006;47(3-4):395-422.
  38. Visscher PM, Hemani G, Vinkhuyzen AAE, et al. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* 2014;10(4):e1004269.
  39. Keller MC, Miller G. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci.* 2006;29(4):385-404; discussion 405-452.
  40. Pappa I, Fedko IO, Mileva-Seitz VR, et al. Single Nucleotide Polymorphism Heritability of Behavior Problems in Childhood: Genome-Wide Complex Trait Analysis. *J Am Acad Child Adolesc Psychiatry.* 2015;54(9):737-744.

## SUPPLEMENTARY METHODS

**Methods 1:** Details related to factor score estimation and imputation

**Methods 2:** Measurement invariance of the internalizing, externalizing and general psychopathology factors across ethnic groups.

**References**



## METHODS 1

### Factor Score estimation

GCTA 1.24.7 only supports observed variables as phenotype. For this reason we estimated the factor scores of the latent factors using the regression method, which were then used for further GREML analyses. A limitation of this approach is, that the relationship between the factors scores may differ slightly from the latent factors. The derived factor scores were based on a confirmatory factor analysis (CFA) fitted in a sample with only complete data ( $n=1954$ ). We established the relationship between the subscales and the general and specific factors and thus were able to predict factor scores using multiple imputation. To assure a reliable prediction, we restricted the estimation of missing factor scores to individuals with up to 50% missing subscales, resulting in a sample of 6,624. The median number of missing scales was 29% with information from multiple informants being available in almost all cases (97%). We based the factor scores on multiply imputed values of the missing subscales. Specifically, we used MICE 2.25 (Buuren & Groothuis-Oudshoorn, 2011) to impute both missing subscales and factor scores, using the subscales, as well as sex and age as predictors in the imputation model. We performed 100 imputations and obtained factor scores by averaging across simulations.

To normalize the distribution of total genetic effects and residuals, as estimated by the best linear unbiased prediction, all factor scores were transformed using  $\ln(\text{score} - \text{lowest score} + 1)$ . This approach of investigating GREML's assumptions, and its limitations, was described in Kirkpatrick, McGue, Iacono, Miller, & Basu (2014)

The total genetic effect for the genetic psychopathology factor was similar between imputed and non-imputed factor scores. We did not use genetic information for the imputation, therefore, if the imputation was of poor quality, we would expect differences in the distribution of the total genetic effect between imputed and non-imputed distributions. We observe a 0.15SD higher score mean for the total genetic effect and slightly narrower standard deviation (0.04SD difference) for children with imputed data as opposed to those without. The similarity in total genetic effect distributions supports successful imputation of the factor score

## METHODS 2

### Measurement invariance of the internalizing, externalizing and general psychopathology factors across ethnic groups

The factor scores were derived from a CFA which included children with any ancestry, we thus implicitly assumed that children scoring equally on the factors would also score equally on the problem subscales regardless whether they have European or non-European ancestry (strong invariance). We tested this assumption formally by

sequentially comparing multi-group CFA models with no to strict equality constraints between European and non-European ancestries (Beaujean, 2014; Hirschfeld & Von Brachel, 2014).

We used a CFI cutoff of  $<0.01$  to decide whether a simpler model with more constraints fits equally well as a more complicated model with less constraints. This fit measure was chosen because it has both been shown to be adequate for testing measurement invariance and because it can be based on a scaled test-statistic when facing non-normal data. It should be noted that performance of detecting measurement invariance in non-normal data has not been examined to the best of our knowledge (Cheung & Rensvold, 2002; Hirschfeld & Von Brachel, 2014). We first present the results for the general psychopathology factor model:

All following general psychopathology factor models are based on 888 children with European ancestry and 437 children without European ancestry, who had complete data on all variables in the general psychopathology factor model as well as genetic information on ancestry. The configural model, which applies the same factor structure to both ancestry groups but no other constraints, had a CFI of 0.890, which is of equivalent magnitude to the originally reported model without groups. Constraining the loadings to be equal between groups (weak invariance) led to an improved fit of 0.896. Additionally constraining the intercepts to be equal (strong invariance) showed a non-significant decrease of the fit to 0.895. The strict fit, which imposes additionally equal residuals, led to a significant decrease in fit to 0.882.

## REFERENCES

- Beaujean, A. (2014). *Latent variable modeling using R: A step-by-step guide* (1st ed.). Routledge.
- Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*.
- Cheung, G., & Rensvold, R. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*.
- Hirschfeld, G., & Von Brachel, R. (2014). Multiple-Group confirmatory factor analysis in R – A tutorial in measurement invariance with continuous and ordinal indicators. *Practical Assessment, Research & Evaluation*, 19(7), 1–11.
- Kirkpatrick, R. M., McGue, M., Iacono, W. G., Miller, M. B., & Basu, S. (2014). Results of a “GWAS Plus:” General Cognitive Ability Is Substantially Heritable and Massively Polygenic. *PLoS ONE*, 9(11), e112390. doi:10.1371/journal.pone.0112390

**Table S1:** Standardized factor loadings of a model with a general psychopathology factor and single combined dimensional internalizing/externalizing factor.

Items	General		Int/Ext Dimension	
	Loading	SE	Loading	SE
Emotional Reactivity (CBCL)	0.61***	0.06	0.01	0.03
Anxious Depressed (CBCL)	0.49***	0.05	-0.14***	0.03
Somatic Complaints (CBCL)	0.25***	0.04	-0.05	0.03
Withdrawn (CBCL)	0.54***	0.07	-0.04	0.03
Attention Problems (CBCL)	0.57***	0.06	0.35***	0.05
Aggressive Behavior (CBCL)	0.62***	0.08	0.22***	0.04
Sleep Problems (CBCL)	0.27***	0.05		
Other (CBCL)	0.61***	0.05		
Social Cognition (SRS)	0.31***	0.04		
Social Communication (SRS)	0.59***	0.07		
Social Autistic Mannerisms (SRS)	0.54***	0.07		
ODD (CPRS-R)	0.44***	0.06	0.14**	0.05
Inattentive (CPRS-R)	0.41***	0.04	0.27***	0.05
Hyperactive-Impulsive (CPRS-R)	0.49***	0.06	0.34***	0.06
Anxious Depressed (TRF)	0.30***	0.06	-0.36***	0.04

**Table S1:** (Continued)

Items	General		Int/Ext Dimension	
	Loading	SE	Loading	SE
Withdrawn Depressed (TRF)	0.31***	0.07	-0.40***	0.04
Somatic Complaints (TRF)	0.16**	0.05	-0.21***	0.04
Social Problems (TRF)	0.36***	0.05		
Thought Problems (TRF)	0.39***	0.07		
Attention Problems (TRF)	0.27***	0.05	0.33***	0.04
Rule-Breaking Behavior (TRF)	0.13**	0.04	0.32***	0.06
Aggressive Behavior (TRF)	0.17***	0.04	0.44***	0.04
Depression (BPI)	0.13***	0.03	0.03	0.04
Seperation Anxiety (BPI)	0.08**	0.03	0.05	0.03
Overanxiousness (BPI)	0.12***	0.03	0.06	0.03
Oppositionality and Defiance (BPI)	0.10**	0.03	0.11**	0.03
Hostility (BPI)	0.11**	0.03	0.18***	0.04
Conduct Problems (BPI)	0.08**	0.03	0.11***	0.03

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

$n = 1954$

**Table S2:** Model fit indices of a model with a general psychopathology factor and single combined dimensional internalizing/externalizing factor.

Model	CFI	TLI	RMSEA <sup>1</sup>	SRMR	BIC
Int/Ext Dimension	0.891	0.856	0.048 [0.046, 0.049]	0.036	283885

<sup>1</sup>A 90% confidence interval is given for RMSEA

n=1954

**Table S3:** Standardized path coefficients from temperament to general psychopathology factor and a combined dimensional internalizing/externalizing factor in a structural equation model.

Temperament	General		Int/Ext Dimension	
	Coefficient	SE	Coefficient	SE
Negative Affectivity	0.53***	0.06	0.06	0.07
Surgency	-0.02	0.06	0.55***	0.04
Effortful Control	-0.22***	0.04	-0.16***	0.03

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

n=1933

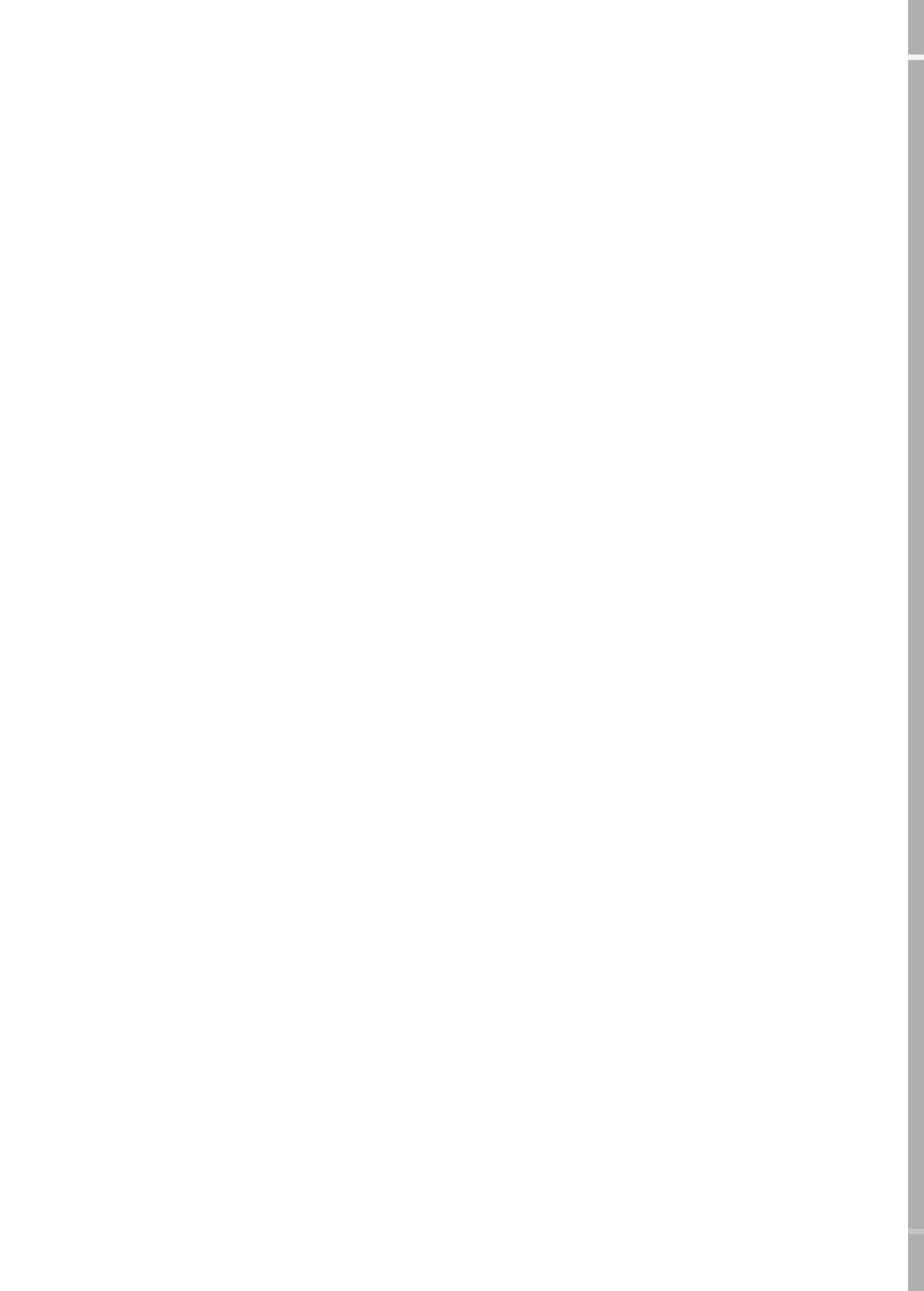






Chapter V

**Cortisol Genetics**





## Chapter V.A

### **Predicting hair cortisol levels with hair pigmentation genes: a possible hair pigmentation bias**

Alexander Neumann; Gerard Noppe; Fan  
Liu; Manfred Kayser; Frank C. Verhulst; Vin-  
cent W.V. Jaddoe, Elisabeth F.C. van Rossum;  
Henning Tiemeier

## ABSTRACT

Cortisol concentrations in hair are used to create hormone profiles spanning months. This method allows assessment of chronic cortisol exposure, but might be biased by hair pigmentation: dark hair was previously related to higher concentrations. It is unclear whether this association arises from local effects, such as increased hormone extractability, or whether the association represents systemic differences arising from population stratification. We tested the hypothesis that hair pigmentation gene variants are associated with varying cortisol levels independent of genetic ancestry. Hormone concentrations and genotype were measured in 1674 children from the Generation R cohort at age 6. We computed a polygenic score of hair color based on 9 single nucleotide polymorphisms. This score was used to predict hair cortisol concentrations, adjusted for genetic ancestry, sex, age and corticosteroid use. A 1-standard deviation (SD) higher polygenic score (darker hair) was associated with 0.08SD higher cortisol levels (SE=0.03,  $p=0.002$ ). This suggests that variation in hair cortisol concentrations is partly explained by local hair effects. In multi-ancestry studies this hair pigmentation bias can reduce power and confound results. Researchers should therefore consider adjusting analyses by reported hair color, by polygenic scores, or by both.

## INTRODUCTION

In the last decade, studies demonstrated that hair is a useful medium to measure chronic cortisol secretion over a period of 3-6 months.<sup>1-4</sup> Each cm of proximal scalp hair represents ca. 1 month cortisol exposure, which makes the measurement of relatively long-term profiles of cortisol and cortisone, a metabolite and precursor of cortisol, feasible.<sup>5</sup> Hair cortisol assessment is therefore an attractive addition to repeated plasma or saliva measurements.

While hair samples are a compelling method, there is concern that hair color might bias measurements. We reported previously that hair color was associated with cortisol and cortisone levels in the Generation R Study, specifically that higher cortisol levels were found in darker hair.<sup>2</sup> Hair pigmentation might directly affect the potential to extract cortisol from hair and thus measured differences may mirror local effects only. Second, the hair cortisol differences might reflect genetic differences in subpopulations (population stratification). For example, since hair color is strongly linked to genetic ancestry, it might be a marker of genetic variations related to cortisol metabolism or sensitivity.<sup>3</sup> Third, color might be a marker for minority status and the related stress, which would explain higher systemic cortisol levels.<sup>1,2,4</sup> Distinguishing between these scenarios is important for observational hair cortisol research, since an association between hair color and hair cortisol might introduce a confounding bias.

To explore the nature of the hair color and hair cortisol association, we investigated whether single nucleotide polymorphisms (SNPs) associated with hair pigmentation are associated with hair cortisol levels in childhood independent of genetic ancestry. For this purpose we selected 16 SNPs from 10 genes included in the HRisPlex system previously developed to predict hair and eye color from DNA.<sup>6,7</sup> We created a polygenic score of hair color, which predicts hair lightness/darkness on the basis of these pigmentation SNPs. We computed a genetic score as opposed to solely using reported hair color, because the score allows a continuous assessment of hair pigmentation, is objective, and is not affected by the environment. Such a genetic score potentially represents hair pigmentation more accurately. This way we tested the main hypothesis that a genetic score of hair color is associated with hair cortisol and cortisone levels independent of genetic ancestry in children.

## METHODS

### Participants

This study was conducted in Generation R. Generation R is a population-based birth cohort aiming to identify early environmental and genetic determinants of development and health.<sup>8,9</sup> All parents gave informed consent for their children's participation.

The Generation R Study is conducted in accordance with the World Medical Association Declaration of Helsinki and study protocols have been approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam.

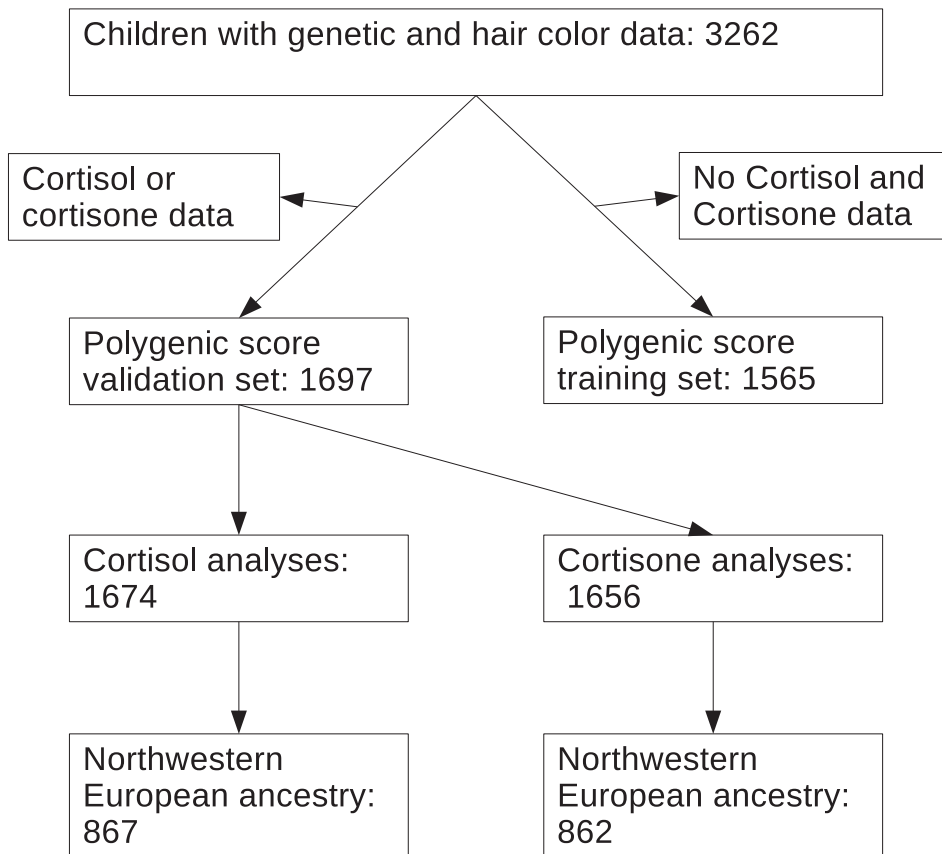
Hair color and genetic information was available in 3262 children. To avoid overfitting of the polygenic score, we split the sample into children with cortisol or cortisone information (n=1697), the validation sample, and a training sample with neither cortisol nor cortisone (n=1565) information. Selection and weights of the SNPs for the polygenic pigmentation score were determined in the training set. Hair cortisol or cortisone measurements were available for 1697 children (1674 had cortisol and 1656 had cortisone available). See Figure 1 for a participant flow chart. Both training and validation samples featured highly admixed populations with a variety of hair phenotypes. See Table 1 for participant characteristics. We additionally studied a subsample of children with genetically northwestern European ancestry to explore whether a hair color bias is present in genetically homogeneous samples. In this sample, 867 measurements of cortisol and 862 of cortisone were available. Finally, we also analyzed subgroups of ethnic minorities grouped by national original of a geographical region: Africa (Africa, Cape Verde, Morocco; n=193), Asia (Asia, Indonesia; n=46), Caribbean (Netherlands Antilles, Suriname; n=156) and Turkey (n=147).

## Genotyping

In Generation R DNA was extracted from whole blood at birth and analyzed using Illumina 610K/660W. We filtered for sample ( $\geq 97.5\%$ ) and SNP call rates ( $\geq 95\%$ ), minor allele frequency  $\geq 1\%$  and deviations from Hardy-Weinberg equilibrium ( $p < 10^{-7}$ ). Excess heterozygosity, gender accuracy, and relatedness were tested. We used MACH 1.0<sup>10</sup> to impute to the 1000 Genomes v3 reference.<sup>11</sup>

We selected 22 SNPs from the HIrisPlex System related to hair color prediction. Nine SNPs were directly genotyped in Generation R and 13 were available as imputed genotypes. Of these, 3 were excluded due to poor imputation quality ( $R^2 < 0.3$ ) and 2 due to a minor allele frequency below 1% (Supplementary Table S1). SNPs were included as allele dosage in all analyses.

Multidimensional scaling was used for the investigation of genetic ancestry based on the genome-wide SNP data.<sup>12</sup> Twenty principal components of ancestry (PCA) were calculated for the whole Generation R sample (n=5731) and subsequently used in the subsample of children with available hair color and hormones data, the training and validation samples. Participants exceeding 4 SDs difference with the mean European reference level (HapMap CEU) on any of the first four principal components were classified as non-northwestern European. For analyses restricted to children with northwestern European ancestry, the PCA were recalculated in that Generation R subsample (n=2830). Again the whole Generation R sample was used for the estimation of PCA. Figure 2 graphically displays the very high population admixture of the training and



**Figure 1:** Participant flow chart

validation samples, by comparing the genetic ancestry to the 1000 Genomes Phase 3 populations.

### Hair Color

Hair color of the children was obtained by parent report and when not available, scored with photographs and videos taken during the research center visit. Inter-coder reliability was calculated with 50 overlapping observations using Krippendorff's alpha. Alpha was 0.79 between the investigators and 0.69 between the investigators and parents.<sup>2</sup> Hair color was categorized into 7 categories: "sandy red" (1), "red or chestnut" (2), "blond" (3), "dark blond" (4), "brown" (5), "dark brown" (6), "brownish black or black" (7), analyzed as continuous variable (ranging from 1-7) indicating pigmentation intensity. See Table 1 for hair color distribution per sample.

**Table 1:** Participant Characteristics

Characteristic	Multi-ancestry			Northwestern European ancestry	
	Training Sample	Cortisol Sample	Cortisone Sample	Cortisol Sample	Cortisone Sample
n	1565	1674	1656	867	862
Cortisol/Cortisone quantiles					
(in pg/mg)					
25%	-	0.91	5.31	0.69	4.80
50%	-	1.65	7.78	1.28	6.61
75%	-	3.26	12.49	2.71	11.85
Hair Color (in %)					
Sandy red	1	2	2	3	3
Red or Chestnut	1	1	1	2	2
Blond	30	26	26	47	46
Dark Blond	33	26	27	38	39
Brown	13	19	18	9	9
Dark Brown	10	15	15	1	1
Brownish black/Black	13	11	10	0	0
National origin (in %)					
Dutch	65	58	59	90	90
Turkish	6	9	9	0	0
Surinamese	6	7	7	0	0
Other European	8	7	7	6	6
Moroccan	3	7	7	0	0

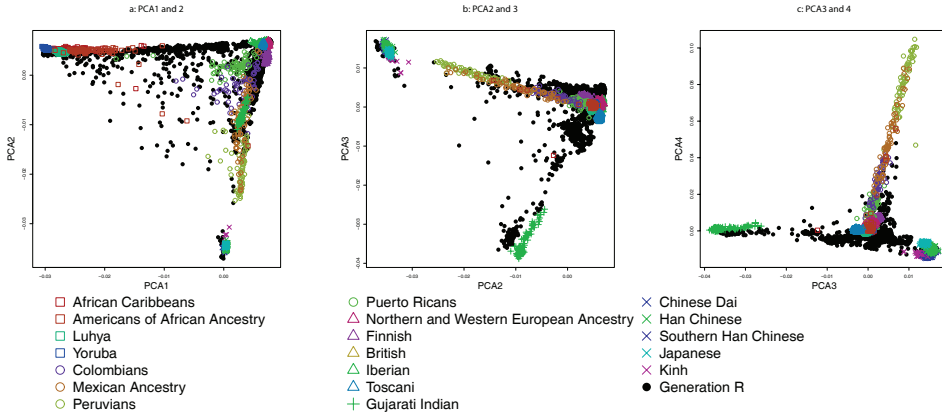


**Table 1:** (Continued)

Cape Verde	2	3	3	0	0
Netherlands Antilles	2	2	2	0	0
African	2	2	2	1	1
American, Non-Western	2	2	2	1	1
Asian, Non-Western	3	2	2	0	0
Indonesian	0	1	1	0	0
Girls (in %)	49	51	52	48	49
Age (mean in years)	6.15	6.19	6.19	6.04	6.04
Corticosteroid use (in %)	-	8	8	9	9

---

## Population Structure in Generation R Study Sample



**Figure 2 a-c:** Comparison of genetic ancestry in the Generation R Study sample and the 1000 Genomes phase 3 populations based on the first four principal components of ancestry (PCA). Squares mark African, circles Ad Mixed American, triangles European, crosses South Asian and X indicates East Asian ancestry.

## Hair Cortisol and Cortisone

Cortisol and cortisone concentrations were measured in the proximal three cm scalp hair, as described previously.<sup>5</sup> Briefly, steroids were extracted using LC-grade methanol at 25°C for 18h in the presence of deuterium labeled steroids as internal standard. Samples were centrifuged and cleaned using solid phase extraction, after which steroids were quantified by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Waters XEVO-TQ-S system, Waters Corporation, Milford, MA, USA), using positive electrospray ionization. See Table 1 for hormone concentrations.

## Statistical Analysis

To determine weights for the hair color polygenic score, we first regressed hair color on HlrisPlex SNPs in a single linear model using the training set ( $n=1565$ ).  $rs16891982$  was not included in the training model due to high multicollinearity (Variance Inflation Factor (VIF)=16.5) caused by strong linkage disequilibrium with  $rs28777$  ( $r^2=0.92$ ). The model was adjusted for 20 PCA ensuring that only SNPs are selected which have explanatory power beyond being markers for genetic ancestry. Using the regression coefficients of the 9 nominally significant ( $\alpha=0.05$ ) SNPs ( $rs885479$ ,  $rs1805008$ ,  $rs1805007$ ,  $rs28777$ ,  $rs12896399$ ,  $rs1042602$ ,  $rs1393350$ ,  $rs12821256$ ,  $rs12203592$ ; see supplementary Table S2) as weights, we calculated a polygenic score in the cortisol/cortisone sam-

ple according to  $\beta_1 * rs885479 + \dots + \beta_8 * rs2378249$ . This resulted in a single score indicating the darkness of the hair.

We used the polygenic score to predict cortisol and cortisone in a linear regression model. We included 20 PCA in the main model as covariates (standardized), next to sex, age and corticosteroid use (parent-reported yes/no). Two to ten PCA are commonly recommended to correct for population stratification, depending on the trait and ancestry admixture.<sup>13</sup> A previous Generation R study suggests that the use of four principal components effectively corrects for population stratification in a genome-wide association study of red hair pigmentation. The performance is comparable to adjustment by linear mixed models.<sup>12</sup> Given the strong correlation of hair color with ancestry, we chose to err on the conservative side and included all 20 PCA.

The ancestry corrected analysis using a polygenic score of hair pigmentation was used to test the main hypothesis. However, we performed additional analyses for exploratory and comparative purposes. We also tested individual hair pigmentation SNPs in separate and in a mutually adjusted model. Next, we related observed hair color (treated continuously) to cortisol and cortisone. As the polygenic score was calculated correcting for ancestry, all models were rerun without the additional genetic ancestry adjustments in the regression analyses. Furthermore, we performed sensitivity analyses in the subsample of children of European ancestry as defined by genetic data. We calculated the power of this subsample to detect effect sizes found in the multi-ancestry sample. For this purpose we used the local Cohen's F2 of the fully adjusted pigmentation effect.<sup>14</sup> Finally, we also stratified the main analysis by four ethnic subgroups (African, Asian, Caribbean, Turkish). Given that these classifications are based on national origin rather than genetic data and that the sample sizes are low, we interpret these analyses exploratory.

The hair cortisol and cortisone regression analyses yielded skewed residual distributions. We therefore applied box-cox transformations. The best fitting lambda was -0.26 for cortisol and -0.06 for cortisone, based on the main model. Transformed values were multiplied with -1 to keep directionality. The polygenic score, cortisol and cortisone were standardized to facilitate interpretation.

To investigate potential pleiotropic effects of the hair pigmentation SNPs, i.e. whether the SNPs are associated with hair cortisol via pathways unrelated to hair color, we investigated the heterogeneity of the single SNP estimates as described by Burgess et al.<sup>15</sup> Estimates and standard errors (SE) were extracted for 9 SNPs from ancestry-adjusted models. We meta-analyzed using inverse-variance weighting to calculate Q and I2 statistics after orienting the SNP effects. A significant Q or high I2 indicate heterogeneity in the effect estimates and can be an indication that the SNP associations are not solely explained by hair color. Heterogeneity is problematic if the pleiotropic effects are in the

same direction for a majority of SNPs, which can be detected visually as asymmetry in a funnel plot or by a significant asymmetry test.<sup>16</sup>

Statistical analyses were performed in R 3.3.2.<sup>17</sup> The package MASS 7.3-45<sup>18</sup> was used for box-cox transformations, metafor 1.9-9<sup>19</sup> for heterogeneity analyses and funnel plots, pwr<sup>20</sup> for power analysis, psych 1.6.9<sup>21</sup> for descriptives and foreign 0.8-67<sup>22</sup> for reading external files.

## RESULTS

In the training set 9 of the 16 SNPs showed nominally significant ( $\alpha=0.05$ ) associations with hair color independent of genetic ancestry as expected (Supplementary Table S2). The polygenic score of the nominally significant SNPs explained 35% of the hair color variance in the validation set ( $n=1697$ ) (adjusted  $R^2$ ). An increase in 1-standard deviation (SD) of the polygenic score was associated with a 0.83-level darker hair ( $\beta=0.83$ ,  $SE=0.03$ ,  $p=8E-160$ ). For comparison, a polygenic score based on 13 SNPs from a model without genetic ancestry adjustment explained 49% of the variance in this sample.

Hair color, individual hair color SNPs, and the polygenic score of hair color predicted hair cortisol and cortisone levels in models unadjusted for genetic ancestry. These models overestimate the effects due to population stratification and are presented solely for comparison with the main analysis. A 1-level darker hair color was associated with 0.16SD higher cortisol levels ( $SE=0.02$ ,  $p=4E-19$ ) and 0.06SD higher cortisone levels ( $SE=0.02$ ,  $p=5E-04$ ) (Table 2 and S3). Six hair color SNPs showed independent nominally significant ( $\alpha=0.05$ ) associations with cortisol and 2 SNPs did with cortisone. (Table 3 and S4). A SD higher polygenic score (darker hair) was associated with 0.21SD higher cortisol levels ( $SE=0.02$ ,  $p=8E-18$ ) and 0.09SD higher cortisone levels ( $SE=0.02$ ,  $p=3E-04$ ) (Table 2 and S3).

The polygenic score explained 4.2% of the hair cortisol variance and 0.7% of hair cortisone in a simple regression (genetic ancestry adjusted in training step), whereas hair color explained 4.5% and 0.6% respectively (no genetic ancestry adjustment). For comparison, a polygenic score based on 13 SNPs from a training model without genetic ancestry adjustment explained 5.8% and 0.8% variance. Genetic ancestry explained 8.0% of the cortisol and 3.8% of the cortisone variance. In contrast, national origin (dummy coded) explained 6.1% and 3.5% respectively.

Introducing genetic ancestry into the models substantially decreased the associations. Darker hair color was not associated with hair cortisol ( $\beta=0.01$ ,  $SE=0.03$ ,  $p=0.70$ ) and cortisone ( $\beta=0.02$ ,  $SE=0.03$ ,  $p=0.46$ ) (Table 2 and S3). However, 2 hair color SNPs remained nominally significant ( $\alpha=0.05$ ) in the cortisol and cortisone models (Table 3 and S4). In the model used for testing the main hypothesis, the polygenic score remained associated with cortisol ( $\beta=0.08$ ,  $SE=0.03$ ,  $p=2E-03$ ) and cortisone ( $\beta=0.06$ ,  $SE=0.03$ ,  $p=0.03$ ) (Table 2 and S3). These models showed no substantial multicollinearity

**Table 2:** Hair cortisol regressed on hair color and polygenic score of hair color in multi-ancestry sample (n=1674). Positive coefficients indicate increases in hormone concentrations. Higher hair color and polygenic scores indicate darker hair. All models were adjusted for sex, age (in months) and corticosteroid (CS) use. Results are shown for models without and with ancestry correction using 20 principal components as covariates (PCA).

Outcome	Hair Cortisol (standardized)											
	Hair Color (no ancestry correction)			Polygenic Score (no ancestry correction)			Hair Color (ancestry correction)			Polygenic Score (ancestry correction)		
Model	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
Intercept	-0.74	0.24	2,00E-3	-0.19	0.24	4,00E-1	0.29	0.27	3,00E-1	0.33	0.24	2,00E-1
Hair Color	0.16	0.02	4,00E-19				0.01	0.03	7,00E-1			
Polygenic score <sup>1</sup>				0.21	0.02	8,00E-18				0.08	0.03	2,00E-3
Sex, female	-0.15	0.05	2,00E-3	-0.14	0.05	4,00E-3	-0.19	0.05	7,00E-5	-0.19	0.05	6,00E-5
Age, months	0.00	0.00	7,00E-1	0.00	0.00	3,00E-1	0.00	0.00	4,00E-1	0.00	0.00	4,00E-1
CS use	0.28	0.09	2,00E-3	0.28	0.09	2,00E-3	0.32	0.09	2,00E-4	0.32	0.09	2,00E-4
PCA1							-0.34	0.04	5,00E-16	-0.30	0.04	8,00E-17
PCA2							-0.04	0.03	1,00E-1	-0.03	0.02	2,00E-1
PCA3							0.08	0.03	2,00E-3	0.07	0.02	4,00E-3
PCA4							0.03	0.02	2,00E-1	0.03	0.02	2,00E-1
PCA5							-0.05	0.02	4,00E-2	-0.05	0.02	3,00E-2
PCA6							-0.03	0.02	2,00E-1	-0.03	0.02	2,00E-1
PCA7							-0.03	0.03	3,00E-1	-0.03	0.03	3,00E-1
PCA8							0.00	0.02	1,00E+0	0.00	0.02	1,00E+0
PCA9							-0.04	0.02	1,00E-1	-0.04	0.02	9,00E-2
PCA10							-0.01	0.03	8,00E-1	0.00	0.03	9,00E-1
PCA11							-0.02	0.02	4,00E-1	-0.02	0.02	4,00E-1
PCA12							-0.03	0.02	3,00E-1	-0.03	0.02	3,00E-1
PCA13							-0.02	0.02	3,00E-1	-0.02	0.02	3,00E-1
PCA14							-0.05	0.02	3,00E-2	-0.05	0.02	3,00E-2

**Table 2:** (Continued)

PCA15	0.04	0.02	1,00E-1	0.04	0.02	1,00E-1
PCA16	0.03	0.02	3,00E-1	0.03	0.02	3,00E-1
PCA17	0.02	0.02	5,00E-1	0.02	0.02	4,00E-1
PCA18	0.00	0.02	1,00E+0	0.00	0.02	1,00E+0
PCA19	-0.05	0.02	4,00E-2	-0.05	0.02	4,00E-2
PCA20	0.00	0.02	1,00E+0	0.00	0.02	1,00E+0

<sup>1</sup>Polygenic score is based on 9 SNPs from a training model adjusted for genetic ancestry

**Table 3:** Hair cortisol regressed on individual pigmentation SNPs in multi-ancestry sample (n=1674).

Predictor	Separate models (no ancestry correction)			Mutually Adjusted (no ancestry correction)			Separate models (ancestry correction)			Mutually Adjusted (ancestry correction)		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
rs885479	-0.04	0.07	6,00E-1	0.04	0.07	6,00E-1	-0.01	0.07	9,00E-1	-0.02	0.08	8,00E-1
rs1805008	0.23	0.07	1,00E-3	0.19	0.07	1,00E-2	0.14	0.07	5,00E-2	0.13	0.07	7,00E-2
rs1805005	0.02	0.08	8,00E-1	0.01	0.08	9,00E-1	-0.10	0.08	2,00E-1	-0.06	0.08	4,00E-1
rs1805007	0.19	0.09	3,00E-2	0.11	0.09	2,00E-1	0.04	0.08	6,00E-1	0.05	0.09	6,00E-1
rs2228479	0.03	0.07	6,00E-1	0.05	0.07	5,00E-1	0.00	0.07	1,00E+0	0.00	0.07	1,00E+0
rs28777	-0.40	0.05	8,00E-18	0.02	0.15	9,00E-1	-0.10	0.06	9,00E-2	-0.01	0.15	1,00E+0
rs16891982	0.39	0.04	7,00E-20	0.31	0.14	3,00E-2	0.11	0.06	8,00E-2	0.10	0.15	5,00E-1
rs2402130	-0.11	0.04	7,00E-3	-0.02	0.04	7,00E-1	-0.01	0.04	7,00E-1	-0.02	0.05	7,00E-1
rs12896399	0.13	0.03	2,00E-4	0.03	0.04	4,00E-1	-0.01	0.04	8,00E-1	-0.02	0.04	7,00E-1
rs1042602	0.03	0.04	4,00E-1	0.01	0.04	7,00E-1	-0.04	0.04	3,00E-1	-0.02	0.04	6,00E-1
rs1393350	0.18	0.04	3,00E-5	0.11	0.05	2,00E-2	0.08	0.04	6,00E-2	0.07	0.05	1,00E-1
rs12821256	0.19	0.06	2,00E-3	0.07	0.06	2,00E-1	0.02	0.06	7,00E-1	0.02	0.06	7,00E-1
rs4959270	0.05	0.04	1,00E-1	0.04	0.03	3,00E-1	0.01	0.03	8,00E-1	0.02	0.03	5,00E-1
rs12203592	-0.07	0.07	3,00E-1	-0.15	0.07	3,00E-2	-0.16	0.06	2,00E-2	-0.17	0.07	1,00E-2
rs1800407	0.01	0.12	9,00E-1	0.07	0.11	6,00E-1	0.04	0.12	7,00E-1	0.05	0.12	6,00E-1
rs2378249	0.11	0.05	2,00E-2	0.13	0.04	5,00E-3	0.10	0.05	3,00E-2	0.11	0.05	2,00E-2
rs683	0.20	0.03	7,00E-10	0.12	0.03	3,00E-4	0.05	0.04	1,00E-1	0.06	0.04	8,00E-2

SNPs were either included in separate models or mutually adjusted in a single model. Positive coefficients indicate increases in hormone concentrations per effect allele (see Table S1, available online).



(all VIF <1.46). Restricting the analysis to children with European ancestry changed the coefficients to 0.05SD for cortisol (SE=0.03,  $p=0.13$ ) and 0.03SD for cortisone (SE=0.03,  $p=0.39$ ) (Table S5-S8), which were not statistically significant. The cortisol analysis in the European subsample had a power of 59% to detect an association of the same magnitude as found in the multi-ancestry sample ( $f^2=0.006$ , power=86%). Repeating the analysis within ethnic minorities revealed a significant association of the polygenic score with hair cortisol in children of African national origin ( $\beta=.22$ , SE=0.09,  $p=0.01$ ) (Table S9). Similar effect sizes in the smaller Caribbean, Asian and Turkish subpopulations did not reach significance.

The associations between 9 single SNPs and hair cortisol showed modest heterogeneity, which was not significant ( $I^2=43.2\%$ ,  $Q=13.6$ ,  $p=0.09$ ). The funnel plot showed no asymmetry (see supplementary Figure S11) and a regression test was not significant ( $p=0.09$ ).

## DISCUSSION

Nine hair color SNPs of the HirisPlex system explained a large proportion of phenotypic hair color variance in the Generation R Study. The polygenic score of hair color was significantly associated with hair cortisol and cortisone levels after strict adjustment for genetic ancestry. The score itself was based only on SNPs, which showed associations with hair color independent of ancestry. The results suggest that cortisol and cortisone levels found in hair are partly explained by hair pigmentation, and do not represent systemic hormone levels only.

Furthermore, the polygenic score of hair color accounted for the variance in the hair hormone concentrations better than parent-reported/photograph-assessed hair color. While the reported hair color did not show associations independent of genetic ancestry, an independent contribution was found for the genetic markers. This suggests that the predictive value of categorical reporting of hair by parents or researchers is lower than the predictive value of the continuous polygenic score. This may seem surprising, given that the polygenic score explained only part of the reported hair color variance. However, the reported hair color was merely used for weighting and for determining the direction of the SNPs. The additional information on allele dosage and number of pigmentation increasing variants is retained and the initial selection of SNP for the HirisPlex system was performed in a separate study. It may therefore well be, that the polygenic score is a better representation of hair pigmentation as opposed to the momentary and subjective hair color report. It should be noted that the performance of the presented polygenic score might change in older children or adults. Hair color can



change with age, thus it is unclear how predictive the presented score is at other ages, since it is calibrated to school aged children.

An association between hair color and hair cortisol levels had been found previously in dogs<sup>23</sup>, as well as in humans<sup>24</sup>, although not in all studies.<sup>25,26</sup> The null results in some previous studies could be due to more homogeneous samples compared to this study, which featured a large number of light and dark haired children. The effects of hair pigmentation on hair cortisol were negligible in the European ancestry subsample, in which dark brown and black hair was virtually absent. This suggests that in samples with lower hair color variation and low ancestry variance the hair color bias on cortisol/cortisone measures may be ignored. However, the power in the European sample was also smaller than in the multi-ancestry sample, which may have limited our ability to detect an association. Furthermore, the hair pigmentation bias remained in the non-European subgroups of children stratified by geography.

At present it is unknown what the exact mechanism is underlying the relation between hair pigmentation and cortisol level measures. However, photocrosslinking between the corticosteroid flumethasone and the protein spectrin has been reported.<sup>27</sup> It is conceivable that dark hair is differently affected by UV radiation than light hair and that may also influence a potential crosslinking of cortisol and hair matrix and thereby cortisol extractability.

Whatever the underlying reasons for the observed phenomenon are, these findings have several important implications. In genetically heterogeneous samples (i.e. participants with ancestry from European as well as other non-European regions), hair color certainly adds additional variance to hormonal measurements, which can increase standard errors, and thus adjustment for hair color or genetic markers of hair color could be beneficial. In genetically homogeneous European samples, bias introduced by hair color is small, as shown here, and may be ignored.

A hair color bias could occur in observational cortisol studies with predictors or outcomes, which are associated with hair color. Such population stratification by hair color is conceivable in studies of metabolic traits, psychological stress, and cortisol genetics among others.

In studies of psychological stress in Western multi-ethnic populations for example, the scenario is possible that dark hair is associated with minority status and consequently increased stress exposure. The observed effects of stress on hair cortisol, however, would then be inflated as the association represents the effects of stress on systemic levels as well as those of dark hair pigmentation on hair cortisol levels. In contrast, if light hair is related to higher stress exposure, associations would be deflated. Such studies are typically adjusted for ethnicity, however, given that ethnicity assessments are imperfect and will not be able to account for all hair color biases, further adjustments for hair color are likely useful. Specifically, polygenic scores are beneficial given that their association with hair hormones is partly independent of genetic ancestry. One might even consider adjusting for a polygenic score only instead of ethnicity to reduce

chances of overadjustment for true stress effects, though some degree of overadjustment may not be avoidable.

Other research situations, in which hair color might cause misleading results, are future genome-wide association or heritability studies of hair cortisol. These study designs might find genetic effects for hair cortisol, which could be completely driven by hair pigmentation genes and their association with local hair hormone levels. Strategies to counter this phenomenon include the exclusion of hair pigmentation genes before analysis, adjustments for hair cortisol genes in the analysis, or the examination of the linkage disequilibrium between genetic association loci and pigmentation variants after analysis.

We used strict adjustments for genetic ancestry in this study. However, residual confounding by ancestry cannot be completely ruled out and hair pigmentation may remain a marker of ancestry even after controlling for principal components. If this scenario were to explain the observed associations, this study would suggest that adjustments for both principal components and genetic markers of hair color are necessary to correct for population stratification in hair cortisol studies.

The possible implications of this study can be summarized as follows: in genetically heterogeneous study populations hair pigmentation bias can reduce power and lead to confounded associations. Researchers should therefore consider adjusting analyses by (reported) hair color, by polygenic scores or both.

## REFERENCES

1. Stalder, T. et al. Stress-related and basic determinants of hair cortisol in humans: A meta-analysis. *Psychoneuroendocrinology* 77, 261–274 (2017).
2. Rippe, R. C. A. et al. Splitting hair for cortisol? Associations of socio-economic status, ethnicity, hair color, gender and other child characteristics with hair cortisol and cortisone. *Psychoneuroendocrinology* 66, 56–64 (2015).
3. Tucker-Drob, E. M. et al. Genetic influences on hormonal markers of chronic hypothalamic-pituitary-adrenal function in human hair. *Psychol. Med.* (2017). doi:10.1017/S0033291716003068
4. Ursache, A., Merz, E. C., Melvin, S., Meyer, J. & Noble, K. G. Socioeconomic Status, Hair Cortisol and Internalizing Symptoms in Parents and Children. *Psychoneuroendocrinology* (2017). doi:10.1016/j.psyneuen.2017.01.020
5. Noppe, G., de Rijke, Y. B., Dorst, K., van den Akker, E. L. T. & van Rossum, E. F. C. LC-MS/MS based method for long-term steroid profiling in human scalp hair. *Clin. Endocrinol. (Oxf)*. 83, 162–166 (2015).
6. Walsh, S. et al. Developmental validation of the HirisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci. Int. Genet.* 9, 150–161 (2014).
7. Branicki, W. et al. Model-based prediction of human hair color using DNA variants. *Hum. Genet.* 443–454 (2011). doi:10.1007/s00439-010-0939-8
8. Tiemeier, H. et al. The Generation R Study: a review of design, findings to date, and a study of the 5-HTTLPR by environmental interaction from fetal life onward. *J. Am. Acad. Child Adolesc. Psychiatry* 51, 1119–1135.e7 (2012).
9. Jaddoe, V. W. V et al. The Generation R Study: design and cohort update 2012. *Eur. J. Epidemiol.* 27, 739–56 (2012).
10. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* 34, 816–834 (2010).
11. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 526, 68–74 (2015).
12. Medina-Gomez, C. et al. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *Eur. J. Epidemiol.* 30, 317–330 (2015).
13. Price, A. L., N.j.patterson, R.m.plenge, M.e.weinblatt & N.a.shadick. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet* 38, 904–909 (2006).
14. Cohen, J. *Statistical power analysis for the behavioral sciences* Lawrence Earlbaum Associates. Hillsdale, NJ 20–26 (1988).
15. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. *Epidemiology* 28, 30–42 (2017).
16. Sterne, J. A. C., Egger, M., Rothstein, H. R. & Sutton, A. J. Regression methods to detect publication and other bias in meta-analysis. *Publ. bias meta-analysis Prev. Assess. Adjust.* 99–110

- (2005).
17. R Core Team. R: A Language and Environment for Statistical Computing. (2016).
  18. Venables, W. N. & Ripley, B. D. Modern Applied Statistics with S. (Springer, 2002).
  19. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* August 36, (2010).
  20. Champely, S. et al. pwr. Available at: <https://github.com/heliosdrm/pwr>. (Accessed: 7th February 2017)
  21. Revelle, W. psych: Procedures for Psychological, Psychometric, and Personality Research. (2016).
  22. R Core Team. foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ... (2016).
  23. Bennett, A. & Hayssen, V. Measuring cortisol in hair and saliva from dogs: Coat color and pigment differences. *Domest. Anim. Endocrinol.* 39, 171–180 (2010).
  24. Staufenbiel, S. M., Penninx, B. W. J. H., de Rijke, Y. B., van den Akker, E. L. T. & van Rossum, E. F. C. Determinants of hair cortisol and hair cortisone concentrations in adults. *Psychoneuroendocrinology* 60, 182–194 (2015).
  25. Manenschijn, L., Koper, J. W., Lamberts, S. W. J. & Rossum, E. F. C. Van. Evaluation of a method to measure long term cortisol levels. *Steroids* 76, 1032–1036 (2011).
  26. Dettenborn, L., Tietze, a., Kirschbaum, C. & Stalder, T. The assessment of cortisol in human hair: Associations with sociodemographic variables and potential confounders. *Stress* 15, 578–588 (2012).
  27. Miolo, G. et al. Photoactivation of corticosteroids in UVB-exposed skin. *J. Photochem. Photobiol. B Biol.* 103, 35–41 (2011).

## SUPPLEMENTARY TABLES AND FIGURES

### Pigmentations SNPs and Hair Color

**Table S1:** HirisPlex SNPs available in Generation R.

SNP	Gene	Effect Allele	Other Allele	Effect Allele Frequency	Minor Allele Frequency	Genotyped	R <sup>2</sup>
rs1042602	TYR	C	A	0.668	0.332	Yes	1.000
rs1110400*	MC1R	T	C	0.995	0.005	No	0.896
rs11547464*	MC1R	G	A	0.995	0.005	No	0.195
rs12203592	IRF4	C	T	0.935	0.065	Yes	1.000
rs12821256	KITLG	T	C	0.916	0.084	Yes	0.978
rs12896399	SLC24A4	G	T	0.614	0.386	Yes	1.000
rs12913832*	HERC2	A	G	0.570	0.430	No	0.208
rs1393350	TYR	G	A	0.824	0.176	Yes	1.000
rs16891982	SLC45A2	C	G	0.262	0.262	No	0.725
rs1800407	OCA2	C	T	0.954	0.046	No	0.471
rs1805005	MC1R	G	T	0.897	0.103	No	0.460
rs1805006*	MC1R	C	A	0.992	0.008	No	0.193
rs1805007	MC1R	C	T	0.953	0.047	No	0.763
rs1805008	MC1R	C	T	0.929	0.071	No	0.789
rs1805009*	MC1R	G	C	0.992	0.008	No	0.340
rs2228479	MC1R	G	A	0.922	0.078	No	0.910
rs2378249	ASIP	A	G	0.838	0.162	Yes	1.000
rs2402130	SLC24A4	A	G	0.769	0.231	Yes	1.000
rs28777	SLC45A2	A	C	0.774	0.226	No	0.696
rs4959270	EXOC2	C	A	0.582	0.418	Yes	1.000
rs683	TYRP1	C	A	0.474	0.474	No	0.988
rs885479	MC1R	G	A	0.941	0.059	Yes	0.999

\* indicates SNPs excluded from analysis due to poor imputation quality and/or low minor allele frequency.

**Table S2:** Hair Color regressed on pigmentation SNPs in multi-ancestry training sample (n=1565).

Predictor	Coefficient	SE	p
(Intercept)	3.18	0.50	3,00E-10
rs885479	0.16	0.07	3,00E-2
rs1805008	0.31	0.07	4,00E-6
rs1805005	0.15	0.08	6,00E-2
rs1805007	0.61	0.08	4,00E-13
rs2228479	0.02	0.06	8,00E-1
rs28777	-0.35	0.07	2,00E-7
rs2402130	-0.05	0.04	3,00E-1
rs12896399	0.09	0.04	2,00E-2
rs1042602	0.09	0.04	2,00E-2
rs1393350	0.10	0.04	3,00E-2

**Table S2:** (Continued)

Predictor	Coefficient	SE	p
rs12821256	0.20	0.05	2,00E-4
rs4959270	0.02	0.03	5,00E-1
rs12203592	-0.46	0.06	7,00E-14
rs1800407	-0.14	0.11	2,00E-1
rs2378249	0.05	0.04	2,00E-1
rs683	0.03	0.03	3,00E-1
PCA1	-15.68	0.88	2,00E-64
PCA2	-17.90	1.28	7,00E-42
PCA3	23.81	1.94	5,00E-33
PCA4	11.43	2.34	1,00E-6
PCA5	-2.45	3.05	4,00E-1
PCA6	0.63	3.62	9,00E-1
PCA7	18.45	3.83	2,00E-6
PCA8	-16.77	4.39	1,00E-4

**Table S2:** (Continued)

---

<b>Predictor</b>	<b>Coefficient</b>	<b>SE</b>	<b>p</b>
PCA9	6.22	6.23	3,00E-1
PCA10	1.41	5.43	8,00E-1
PCA11	-2.61	5.27	6,00E-1
PCA12	4.90	5.51	4,00E-1
PCA13	1.49	5.02	8,00E-1
PCA14	8.07	5.00	1,00E-1
PCA15	-7.40	4.61	1,00E-1
PCA16	0.46	4.72	9,00E-1
PCA17	-8.72	4.67	6,00E-2
PCA18	5.98	4.63	2,00E-1
PCA19	10.43	4.72	3,00E-2
PCA20	-3.76	4.56	4,00E-1

---

Regression coefficients indicate increase in 1 level darker hair per number of effect allele.



## Hair Cortisone (all ancestries)

**Table S3:** Hair cortisone regressed on hair color and polygenic score of hair color in multi-ancestry sample (n=1656).

Outcome	Hair Cortisone (standardized)															
	Hair Color (no ancestry correction)				Polygenic Score (no ancestry correction)				Hair Color (ancestry correction)				Polygenic Score (ancestry correction)			
	$\beta$	SE	p		$\beta$	SE	p		$\beta$	SE	p		$\beta$	SE	p	
Intercept	-0.22	0.24	4,00E-1		0.01	0.24	1,00E+0		0.07	0.28	8,00E-1		0.16	0.25	5,00E-1	
Hair Color	0.06	0.02	5,00E-4					0.02	0.03	5,00E-1						
Polygenic score <sup>1</sup>				0.09	0.02	3,00E-4							0.06	0.03	3,00E-2	
Sex, female	-0.22	0.05	1,00E-5		-0.21	0.05	1,00E-5		-0.23	0.05	2,00E-6		-0.23	0.05	2,00E-6	
Age, months	0.00	0.00	8,00E-1		0.00	0.00	6,00E-1		0.00	0.00	8,00E-1		0.00	0.00	9,00E-1	
CS use	-0.09	0.09	3,00E-1		-0.09	0.09	3,00E-1									
PCA1								-0.08	0.09	4,00E-1		-0.08	0.09	4,00E-1		
PCA2								0.05	0.04	2,00E-1		0.07	0.04	7,00E-2		
PCA3								-0.07	0.03	1,00E-2		-0.07	0.02	7,00E-3		
PCA4								0.17	0.03	4,00E-10		0.17	0.03	4,00E-11		
PCA5								-0.03	0.02	2,00E-1		-0.03	0.02	2,00E-1		
PCA6								-0.06	0.02	1,00E-2		-0.06	0.02	9,00E-3		
PCA7								-0.06	0.03	3,00E-2		-0.06	0.03	3,00E-2		
PCA8								-0.02	0.03	4,00E-1		-0.02	0.03	4,00E-1		
PCA9								-0.03	0.03	2,00E-1		-0.03	0.03	2,00E-1		
PCA10								-0.03	0.02	2,00E-1		-0.03	0.02	2,00E-1		
PCA11								0.01	0.03	8,00E-1		0.01	0.03	7,00E-1		
PCA12								-0.01	0.02	6,00E-1		-0.01	0.02	5,00E-1		
PCA13								-0.01	0.02	7,00E-1		-0.01	0.02	7,00E-1		
PCA14								0.00	0.02	1,00E+0		0.00	0.02	1,00E+0		
PCA15								-0.06	0.03	2,00E-2		-0.06	0.03	2,00E-2		
PCA16								0.02	0.02	4,00E-1		0.02	0.02	4,00E-1		
								-0.03	0.02	2,00E-1		-0.03	0.02	2,00E-1		

**Table S3:** (Continued)

Outcome Model Predictor	Hair Cortisone (standardized)											
	Hair Color (no ancestry correction)			Polygenic Score (no ancestry correction)			Hair Color (ancestry correction)			Polygenic Score (ancestry correction)		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
PCA17				-0.01	0.02	6,00E-1	-0.01	0.02	6,00E-1	-0.01	0.02	6,00E-1
PCA18				0.00	0.02	1,00E+0	0.00	0.02	1,00E+0	0.00	0.02	1,00E+0
PCA19				-0.02	0.02	3,00E-1	-0.02	0.02	3,00E-1	-0.02	0.02	3,00E-1
PCA20				0.02	0.02	5,00E-1	0.02	0.02	5,00E-1	0.01	0.02	5,00E-1

Positive coefficients indicate increases in hormone concentrations. Higher hair color and polygenic scores indicate darker hair. All models were adjusted for sex, age (in months) and corticosteroid (CS) use. Results are shown for models without and with ancestry correction (PCA).  
 1Polygenic score is based on 9 SNPs from a training model adjusted for genetic ancestry

**Table S4:** Hair cortisone regressed on individual pigmentation SNPs in multi-ancestry sample (n=1656).

Predictor	Separate models (no ancestry correction)			Mutually Adjusted (no ancestry correction)			Separate models (ancestry correction)			Mutually Adjusted (ancestry correction)		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
rs885479	-0.01	0.07	9,00E-1	0.01	0.07	9,00E-1	0.03	0.07	7,00E-1	0.05	0.08	5,00E-1
rs1805008	0.28	0.07	1,00E-4	0.26	0.07	6,00E-4	0.26	0.07	3,00E-4	0.27	0.07	3,00E-4
rs1805005	-0.08	0.08	3,00E-1	-0.05	0.08	5,00E-1	-0.04	0.08	6,00E-1	0.02	0.09	8,00E-1
rs1805007	0.10	0.09	3,00E-1	0.07	0.09	4,00E-1	0.01	0.09	9,00E-1	0.05	0.09	6,00E-1
rs2228479	-0.03	0.07	7,00E-1	-0.03	0.07	7,00E-1	-0.05	0.07	5,00E-1	-0.02	0.07	8,00E-1
rs28777	-0.13	0.05	8,00E-3	0.06	0.16	7,00E-1	-0.04	0.06	5,00E-1	0.08	0.16	6,00E-1
rs16891982	0.13	0.04	4,00E-3	0.11	0.15	4,00E-1	0.06	0.06	3,00E-1	0.14	0.16	4,00E-1
rs2402130	-0.05	0.04	2,00E-1	-0.01	0.05	9,00E-1	-0.06	0.04	2,00E-1	-0.05	0.05	3,00E-1
rs12896399	0.08	0.03	2,00E-2	0.05	0.04	2,00E-1	0.05	0.04	2,00E-1	0.03	0.04	4,00E-1
rs1042602	-0.06	0.04	1,00E-1	-0.06	0.04	1,00E-1	-0.03	0.04	4,00E-1	-0.01	0.04	7,00E-1
rs1393350	0.10	0.04	2,00E-2	0.06	0.05	2,00E-1	0.09	0.04	5,00E-2	0.08	0.05	8,00E-2
rs12821256	0.10	0.06	9,00E-2	0.05	0.06	4,00E-1	0.02	0.06	8,00E-1	0.01	0.06	9,00E-1
rs4959270	-0.02	0.04	5,00E-1	-0.04	0.04	3,00E-1	-0.03	0.03	4,00E-1	-0.03	0.04	4,00E-1
rs12203592	0.03	0.07	7,00E-1	0.02	0.07	8,00E-1	-0.01	0.07	9,00E-1	0.00	0.07	1,00E+0

**Table S4:** (Continued)

Predictor	Separate models (no ancestry correction)			Mutually Adjusted (no ancestry correction)			Separate models (ancestry correction)			Mutually Adjusted (ancestry correction)		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
rs1800407	-0.16	0.12	2,00E-1	-0.14	0.12	2,00E-1	-0.11	0.12	3,00E-1	-0.10	0.12	4,00E-1
rs2378249	0.14	0.05	3,00E-3	0.14	0.05	2,00E-3	0.12	0.05	1,00E-2	0.12	0.05	7,00E-3
rs683	0.06	0.03	6,00E-2	0.03	0.03	3,00E-1	0.02	0.04	6,00E-1	0.03	0.04	5,00E-1

SNPs were either included in separate models or mutually adjusted in a single model. Positive coefficients indicate increases in hormone concentrations per effect allele (see Table S1).

## Cortisol (European Ancestry)

Table S5: Hair cortisol regressed on hair color and polygenic score of hair color in European sample (n=867).

Outcome	Hair Cortisol (standardized, European ancestry)											
	Hair Color (no ancestry correction)			Polygenic Score (no ancestry correction)			Hair Color (ancestry correction)			Polygenic Score (ancestry correction)		
Model	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
Intercept	0.41	0.43	3,00E-1	0.38	0.41	3,00E-1	0.42	0.43	3,00E-1	0.41	0.41	3,00E-1
Hair Color	-0.01	0.04	8,00E-1				-0.01	0.04	9,00E-1			
Polygenic score <sup>1</sup>				0.05	0.03	2,00E-1				0.05	0.03	1,00E-1
Sex, female	-0.26	0.07	1,00E-4	-0.25	0.07	2,00E-4	-0.28	0.07	6,00E-5	-0.27	0.07	7,00E-5
Age, months	0.00	0.01	5,00E-1	0.00	0.01	5,00E-1	0.00	0.01	5,00E-1	0.00	0.01	5,00E-1
CS use	0.35	0.12	3,00E-3	0.35	0.12	3,00E-3	0.35	0.12	3,00E-3	0.36	0.12	2,00E-3
PCA1							-0.01	0.03	8,00E-1	-0.01	0.03	8,00E-1
PCA2							-0.04	0.05	5,00E-1	-0.04	0.05	5,00E-1
PCA3							0.00	0.03	1,00E+0	0.00	0.03	9,00E-1
PCA4							0.07	0.03	4,00E-2	0.07	0.03	5,00E-2
PCA5							-0.03	0.03	3,00E-1	-0.03	0.03	3,00E-1
PCA6							0.03	0.03	4,00E-1	0.03	0.03	3,00E-1
PCA7							-0.02	0.04	6,00E-1	-0.02	0.04	6,00E-1
PCA8							-0.02	0.04	6,00E-1	-0.02	0.04	5,00E-1
PCA9							0.02	0.03	5,00E-1	0.02	0.03	5,00E-1
PCA10							-0.05	0.03	1,00E-1	-0.05	0.03	1,00E-1
PCA11							0.06	0.04	1,00E-1	0.05	0.04	1,00E-1
PCA12							0.04	0.03	2,00E-1	0.04	0.03	2,00E-1
PCA13							0.04	0.03	2,00E-1	0.04	0.03	2,00E-1
PCA14							-0.01	0.03	7,00E-1	-0.01	0.03	7,00E-1
PCA15							0.05	0.03	2,00E-1	0.05	0.03	2,00E-1
PCA16							-0.03	0.03	4,00E-1	-0.03	0.03	3,00E-1

**Table S5:** (Continued)

Outcome Model	Hair Cortisol (standardized, European ancestry)								
	Hair Color (no ancestry correction)		Polygenic Score (no ancestry correction)		Hair Color (ancestry correction)		Polygenic Score (ancestry correction)		
Predictor	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
PCA17				0.02	0.03	5,00E-1	0.02	0.03	5,00E-1
PCA18				0.02	0.03	6,00E-1	0.02	0.03	5,00E-1
PCA19				-0.01	0.03	9,00E-1	-0.01	0.03	9,00E-1
PCA20				0.03	0.03	4,00E-1	0.03	0.03	4,00E-1

Positive coefficients indicate increases in hormone concentrations. Higher hair color and polygenic scores indicate darker hair. All models were adjusted for sex, age (in months) and corticosteroid (CS) use. Results are shown for models without and with ancestry correction (PCA).  
<sup>1</sup>Polygenic score is based on 9 SNPs from a training model adjusted for genetic ancestry

**Table S6:** Hair cortisol regressed on individual pigmentation SNPs in European ancestry sample (n=867).

Predictor	Separate models (no ancestry correction)				Mutually Adjusted (no ancestry correction)				Separate models (ancestry correction)				Mutually Adjusted (ancestry correction)			
	$\beta$	SE	p		$\beta$	SE	p		$\beta$	SE	p		$\beta$	SE	p	
rs885479	-0.13	0.11	2,00E-1		-0.16	0.11	2,00E-1		-0.13	0.11	2,00E-1		-0.16	0.11	1,00E-1	
rs1805008	0.14	0.09	1,00E-1		0.11	0.09	2,00E-1		0.15	0.09	1,00E-1		0.12	0.10	2,00E-1	
rs1805005	-0.15	0.11	2,00E-1		-0.14	0.12	2,00E-1		-0.18	0.11	9,00E-2		-0.18	0.12	1,00E-1	
rs1805007	-0.01	0.10	9,00E-1		0.00	0.11	1,00E+0		0.00	0.10	1,00E+0		0.00	0.11	1,00E+0	
rs2228479	0.10	0.09	3,00E-1		0.07	0.09	5,00E-1		0.10	0.09	3,00E-1		0.07	0.09	5,00E-1	
rs28777	-0.12	0.16	5,00E-1		0.01	0.52	1,00E+0		-0.13	0.16	4,00E-1		0.01	0.52	1,00E+0	
rs16891982	0.10	0.15	5,00E-1		0.11	0.47	8,00E-1		0.12	0.15	4,00E-1		0.11	0.47	8,00E-1	
rs2402130	0.00	0.06	9,00E-1		-0.03	0.07	7,00E-1		0.00	0.06	1,00E+0		-0.04	0.07	5,00E-1	
rs12896399	-0.03	0.05	5,00E-1		-0.04	0.05	4,00E-1		-0.04	0.05	4,00E-1		-0.06	0.05	3,00E-1	
rs1042602	-0.05	0.05	4,00E-1		-0.03	0.06	6,00E-1		-0.06	0.05	2,00E-1		-0.04	0.06	5,00E-1	
rs1393350	0.06	0.06	3,00E-1		0.06	0.06	3,00E-1		0.08	0.06	1,00E-1		0.08	0.06	2,00E-1	
rs12821256	-0.07	0.07	3,00E-1		-0.06	0.07	4,00E-1		-0.06	0.07	4,00E-1		-0.06	0.07	4,00E-1	
rs4959270	-0.01	0.05	8,00E-1		0.01	0.05	8,00E-1		-0.01	0.05	8,00E-1		0.01	0.05	8,00E-1	
rs12203592	-0.21	0.08	1,00E-2		-0.21	0.09	2,00E-2		-0.21	0.09	2,00E-2		-0.20	0.09	2,00E-2	

**Table S6:** (Continued)

Predictor	Separate models (no ancestry correction)			Mutually Adjusted (no ancestry correction)			Separate models (ancestry correction)			Mutually Adjusted (ancestry correction)		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
rs1800407	0.08	0.16	6,00E-1	0.06	0.16	7,00E-1	0.10	0.17	6,00E-1	0.08	0.17	6,00E-1
rs2378249	0.13	0.06	4,00E-2	0.14	0.06	2,00E-2	0.13	0.06	4,00E-2	0.14	0.06	3,00E-2
rs683	0.03	0.05	6,00E-1	0.03	0.05	6,00E-1	0.03	0.05	5,00E-1	0.03	0.05	6,00E-1

SNPs were either included in separate models or mutually adjusted in a single model. Positive coefficients indicate increases in hormone concentrations per effect allele (see Table S1).



## Cortisone (European Ancestry)

Table S7: Hair cortisone regressed on hair color and polygenic score in European sample (n=862).

Outcome	Hair Cortisone (standardized, European ancestry)					
	Hair Color (no ancestry correction)			Polygenic Score (no ancestry correction)		
Model	$\beta$	SE	p	$\beta$	SE	p
Predictor						
Intercept	-0.04	0.43	9,00E-1	0.09	0.41	8,00E-1
Hair Color	0.04	0.04	3,00E-1	-0.02	0.43	1,00E+0
				0.04	0.04	3,00E-1
Polygenic score <sup>a</sup>				0.03	0.03	4,00E-1
Sex, female	-0.30	0.07	1,00E-5	-0.29	0.07	2,00E-5
Age, months	0.00	0.01	9,00E-1	0.00	0.01	9,00E-1
CS use	-0.03	0.12	8,00E-1	-0.03	0.12	8,00E-1
PCA1				-0.03	0.12	8,00E-1
PCA2				0.01	0.03	8,00E-1
PCA3				0.04	0.05	4,00E-1
PCA4				-0.02	0.03	7,00E-1
PCA5				0.02	0.03	5,00E-1
PCA6				0.00	0.03	1,00E+0
PCA7				-0.01	0.03	8,00E-1
PCA8				0.00	0.04	9,00E-1
PCA9				-0.04	0.04	2,00E-1
PCA10				-0.05	0.03	1,00E-1
PCA11				-0.04	0.03	2,00E-1
PCA12				0.01	0.04	9,00E-1
PCA13				0.01	0.03	7,00E-1
PCA14				0.03	0.03	4,00E-1
PCA15				-0.03	0.03	3,00E-1
				-0.01	0.03	8,00E-1
				0.00	0.01	1,00E+0
				-0.02	0.12	8,00E-1
				0.01	0.03	8,00E-1
				0.04	0.05	4,00E-1
				-0.02	0.03	7,00E-1
				0.02	0.03	5,00E-1
				0.00	0.03	1,00E+0
				-0.01	0.03	8,00E-1
				0.00	0.04	9,00E-1
				-0.04	0.04	2,00E-1
				-0.05	0.03	1,00E-1
				-0.04	0.03	2,00E-1
				0.00	0.04	9,00E-1
				0.01	0.03	7,00E-1
				0.03	0.03	4,00E-1
				-0.03	0.03	3,00E-1
				-0.01	0.03	8,00E-1
				0.03	0.03	4,00E-1
				-0.30	0.07	1,00E-5
				0.00	0.01	1,00E+0
				-0.02	0.12	8,00E-1
				0.01	0.03	8,00E-1
				0.04	0.05	4,00E-1
				-0.02	0.03	7,00E-1
				0.02	0.03	5,00E-1
				0.00	0.03	1,00E+0
				-0.01	0.03	8,00E-1
				0.00	0.04	9,00E-1
				-0.04	0.04	2,00E-1
				-0.05	0.03	1,00E-1
				-0.04	0.03	2,00E-1
				0.00	0.04	9,00E-1
				0.01	0.03	7,00E-1
				0.03	0.03	4,00E-1
				-0.03	0.03	3,00E-1
				-0.01	0.03	8,00E-1
				0.12	0.42	8,00E-1

**Table S7:** (Continued)

Outcome Model	Hair Cortisone (standardized, European ancestry)											
	Hair Color (no ancestry correction)			Polygenic Score (no ancestry correction)			Hair Color (ancestry correction)			Polygenic Score (ancestry correction)		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
PCA16												
PCA17												
PCA18												
PCA19												
PCA20												

Positive coefficients indicate increases in hormone concentrations. Higher hair color and polygenic scores indicate darker hair. All models were adjusted for sex, age (in months) and corticosteroid (CS) use. Results are shown for models without and with ancestry correction (PCA).  
 1Polygenic score is based on 9 SNPs from a training model adjusted for genetic ancestry

**Table S8:** Hair cortisone regressed on individual pigmentation SNPs in European ancestry sample (n=862).

Predictor	Separate models (no ancestry correction)			Mutually Adjusted (no ancestry correction)			Separate models (ancestry correction)			Mutually Adjusted (ancestry correction)		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
rs885479	-0.06	0.10	6,00E-1	-0.07	0.11	5,00E-1	-0.05	0.11	6,00E-1	-0.07	0.11	6,00E-1
rs1805008	0.20	0.09	2,00E-2	0.20	0.09	4,00E-2	0.21	0.09	2,00E-2	0.21	0.1	3,00E-2
rs1805005	-0.09	0.11	4,00E-1	-0.07	0.12	6,00E-1	-0.10	0.11	4,00E-1	-0.07	0.12	5,00E-1
rs1805007	-0.02	0.10	8,00E-1	0.00	0.11	1,00E+0	-0.01	0.1	9,00E-1	0.02	0.11	8,00E-1
rs2228479	0.05	0.09	6,00E-1	0.05	0.09	6,00E-1	0.06	0.09	5,00E-1	0.06	0.1	5,00E-1
rs28777	0.01	0.16	9,00E-1	0.81	0.52	1,00E-1	-0.02	0.16	9,00E-1	0.75	0.53	2,00E-1
rs16891982	0.06	0.14	7,00E-1	0.76	0.47	1,00E-1	0.08	0.15	6,00E-1	0.73	0.48	1,00E-1
rs2402130	-0.03	0.06	7,00E-1	-0.05	0.07	5,00E-1	-0.03	0.06	6,00E-1	-0.06	0.07	4,00E-1
rs12896399	0.00	0.05	9,00E-1	-0.02	0.05	8,00E-1	-0.01	0.05	8,00E-1	-0.03	0.06	6,00E-1
rs1042602	-0.02	0.05	8,00E-1	0.02	0.06	7,00E-1	-0.03	0.05	6,00E-1	0.01	0.06	8,00E-1
rs1393350	0.12	0.06	4,00E-2	0.13	0.06	3,00E-2	0.13	0.06	3,00E-2	0.14	0.06	2,00E-2
rs12821256	-0.03	0.07	7,00E-1	-0.04	0.07	6,00E-1	-0.02	0.07	8,00E-1	-0.03	0.07	7,00E-1
rs4959270	0.01	0.05	9,00E-1	0.00	0.05	9,00E-1	0.00	0.05	1,00E+0	0.00	0.05	1,00E+0

**Table S8:** (Continued)

Predictor	Separate models (no ancestry correction)			Mutually Adjusted (no ancestry correction)			Separate models (ancestry correction)			Mutually Adjusted (ancestry correction)		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
rs122203592	-0.01	0.08	9,00E-1	-0.01	0.09	9,00E-1	0.00	0.09	1,00E+0	0.00	0.09	1,00E+0
rs1800407	0.02	0.17	9,00E-1	0.00	0.17	1,00E+0	0.01	0.17	9,00E-1	0.00	0.17	1,00E+0
rs2378249	0.11	0.06	7,00E-2	0.12	0.06	5,00E-2	0.11	0.06	8,00E-2	0.12	0.06	6,00E-2
rs683	0.02	0.05	6,00E-1	0.02	0.05	7,00E-1	0.02	0.05	7,00E-1	0.02	0.05	7,00E-1

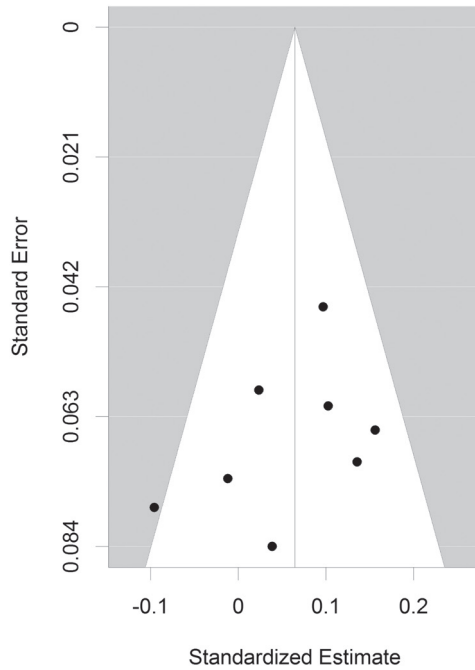
SNPs were either included in separate models or mutually adjusted in a single model. Positive coefficients indicate increases in hormone concentrations per effect allele (see Table S1).

**Table S9:** Hair cortisol and cortisone regressed on polygenic score of hair color stratified by national origin of ethnic minorities

National Origin	n	$\beta$	SE	p
<i>Cortisol</i>				
Africa	193	0.22	0.09	1,00E-2
Asia	46	0.29	0.27	3,00E-1
Caribbean	156	0.08	0.11	5,00E-1
Turkey	147	0.17	0.10	8,00E-2
<i>Cortisone</i>				
Africa	185	0.15	0.09	8,00E-2
Asia	43	0.34	0.22	1,00E-1
Caribbean	153	0.24	0.15	1,00E-1
Turkey	141	0.09	0.11	4,00E-1

Positive coefficients indicate increases in hormone concentrations. Higher polygenic scores indicate darker hair. All models were adjusted for sex, age (in months), corticosteroid use and genetic ancestry.





**Figure S11:** Funnel plot showing standardized estimates of 9 SNPs associated with hair cortisol against their standard error. White area indicates 95% confidence interval. SNPs not included in polygenic score were omitted. Each model was adjusted for sex, age (in months), corticosteroid (CS) use and genetic ancestry (PCA).



## Chapter V.B

### **The Low Single Nucleotide Polymorphism Heritability of Plasma and Saliva Cortisol Lev- els**

Alexander Neumann; Nese Direk; Andrew A Crawford; Saira Mirza; Hieab Adams; Jennifer Bolton; Caroline Hayward; David P Strachan; Erin K. Payne, Jennifer A. Smith; Yuri Milaneschi; Brenda Penninx; Jouke J Hottenga; Eco de Geus; Albertine J Oldehinkel; Peter J van der Most; Yolanda de Rijke; Brian R Walker; Henning Tiemeier

## ABSTRACT

Cortisol is an important stress hormone affected by a variety of biological and environmental factors, such as the circadian rhythm, exercise and psychological stress. Cortisol is mostly measured using blood or saliva samples. A number of genetic variants have been found to contribute to cortisol levels with these methods. While the effects of several specific single genetic variants is known, the joint genome-wide contribution to cortisol levels is unclear. Our aim was to estimate the amount of cortisol variance explained by common single nucleotide polymorphisms, i.e. the SNP heritability, using a variety of cortisol measures, cohorts and analysis approaches. We analyzed morning plasma ( $n=5,705$ ) and saliva levels ( $n=1,717$ ), as well as diurnal saliva levels ( $n=1,541$ ), in the Rotterdam Study using genomic restricted maximum likelihood estimation. Additionally, linkage disequilibrium score regression was fitted on the results of genome-wide association studies (GWAS) performed by the CORNET consortium on morning plasma cortisol ( $n=12,597$ ) and saliva cortisol ( $n=7,703$ ). No significant SNP heritability was detected for any cortisol measure, sample or analysis approach. Point estimates ranged from 0% to 9%. Morning plasma cortisol in the CORNET cohorts, the sample with the most power, had a 6% [95%CI: 0-13%] SNP heritability. The results consistently suggest a low SNP heritability of these acute and short-term measures of cortisol. The low SNP heritability may reflect the substantial environmental and, in particular, situational component of these cortisol measures. Future GWAS will require very large sample sizes. Alternatively, more long-term cortisol measures such as hair cortisol samples are needed to discover further genetic pathways regulating cortisol concentrations.



## INTRODUCTION

Cortisol secretion is regulated by the hypothalamic-pituitary-adrenal axis in response to various biological and environmental factors, including physical stressors such as intensive resistance exercise<sup>1</sup> or injury,<sup>2</sup> and psychological stressors such as public speaking and demanding cognitive tasks.<sup>3</sup> Cortisol secretion has a marked circadian rhythm: secretion peaks shortly after awakening and then drops throughout the day, reflecting the hormone's role in regulating energy metabolism.<sup>4</sup> Additionally, cortisol is secreted rhythmically resulting in a pulsatile ultradian rhythm.<sup>5</sup> The combination of these factors leads to substantial systematic and unsystematic variation of cortisol levels throughout the day.

Cortisol levels can be assessed with a variety of methods, the most common being blood in plasma and saliva samples. Plasma samples represent bound and unbound cortisol concentrations, whereas saliva represents the bioactive free cortisol. These measures have a modest to good correlation<sup>6,7</sup> and have been associated with various traits and states: BMI,<sup>8</sup> cardiovascular risk factors including hyperglycaemia,<sup>9</sup> psychiatric disorders, such as post-traumatic stress disorder, schizophrenia or bipolar disorder<sup>10,11</sup> and treatment response to depression.<sup>12</sup> Saliva cortisol can be sampled non-invasively, which may reduce the chance of inducing stress, makes repeated measurements more feasible, and facilitates mapping of day-time profiles. Repeated cortisol measures tend to show higher between-visit reliability than single measures at awakening or 8am.<sup>13,14</sup>

Plasma and saliva cortisol have been investigated in twin studies to determine the extent of the genetic contribution underlying the hormone. For acute plasma cortisol measures, the estimates range from low (14%) to moderate heritability (45%).<sup>15–17</sup> Wüst et al.<sup>18</sup> reported 0% heritability for acute saliva levels at 8am and total day-time profiles, and observed a large contribution of shared environment (>40%). These family studies rely on relatedness information obtained from known familiar relationships instead of direct molecular measurements such as SNP arrays. Molecular genetic studies that can clarify the nature and extent of the genetic effects underlying cortisol are lacking, although they could advance our understanding of the genetic contribution to stress vulnerability as assessed by cortisol. A genome-wide association study (GWAS) by the cortisol network consortium (CORNET) successfully detected and replicated one genetic locus associated with morning plasma cortisol levels, suggesting that common autosomal gene variants are associated with this phenotype.<sup>19</sup> It is plausible that a substantial number of variants associated with cortisol were not identified due to stringent multiple testing corrections required in GWAS. If this is the case, then the joint effect of all SNPs should be larger than the variance explained by the locus found (<1%).

In the present study, we aimed to quantify the SNP heritability of cortisol, i.e the variance jointly explained by common autosomal single nucleotide polymorphisms. The SNP heritability information represents a more direct measure of the genetic pre-

disposition to high or low cortisol stemming from additive genetic effects of common gene variants compared to the broad-sense heritability estimated in family studies. SNP heritability can therefore inform future GWA studies about sample size and potential success. We focus on cortisol measured in plasma and saliva measured in elderly participants from the Rotterdam Study and in mixed ages from the CORNET cohorts. This allowed the study of acute morning levels (plasma and saliva) and day-time profiles (saliva) in large sample sizes. SNP heritability can be estimated with different methods. In this study we used genomic restricted maximum likelihood estimation (GREML)<sup>20</sup> in the Rotterdam Study as well as LD score regression in the CORNET GWAS results.

## METHODS

### Rotterdam Study

#### Participants

The Rotterdam Study is a population-based cohort investigating chronic disease and their risk factors in elderly, see Hofman et al.<sup>21</sup> for details. The Rotterdam Study includes 14,926 participants aged 45 and older. Study protocols were approved by the medical ethics committee according to the Population Study Act Rotterdam Study, executed by the Ministry of Health, Welfare and Sports of the Netherlands. Written informed consent was obtained from all participants.

Plasma cortisol information was available in 9836 participants performed in 1997-2008. For 8501, complete information on genetics was available. 2796 participants were removed from GREML analyses due to excessive relatedness (see 2.1.2), resulting in a GREML sample of 5705. In the time adjusted analyses, a further 83 were excluded due to missing information regarding timing of sampling.

Saliva cortisol was available in 2034 participants of which 1982 had complete data on genetics. After removal of 265 participants due to excessive relatedness 1717 individuals remained with acute saliva level upon awakening. Of those, 1541 had also information on later time points for total day-time cortisol computations. See Table 1 for participant characteristics.

#### Measurements

Plasma cortisol was collected from 8:00h to 20:00h. 75% of samples were collected before 10:30 and 99% before 15:30. Cortisol was measured using the LC-MS/MS method with the CHS MSMS Steroids Kit (Perkin Elmer, Turku, Finland) containing 2H3-cortisol as internal standard. Chromatographic separation was performed on a Waters

(Milford, MA, USA) Acquity UPLC HSS T3 1.8 $\mu$ m column and quantified by tandem mass spectrometry using a Xevo TQ-S system (Waters, Milford, MA).

Sarstedt Cortisol Salivette collection tubes (Sarstedt, Rommelsdorf, Germany) were used to collect saliva after awakening, 30 min after awakening, at 17:00 and at bedtime by the participants.<sup>22</sup> Participants were instructed to note the exact time of saliva collection, and not to eat or brush teeth 15min before collection. An enzyme immunoassay (IBL International GmbH Hamburg, Hamburg, Germany) was used to analyze the samples. We investigated awakening cortisol levels and diurnal cortisol, calculated by the area under the curve in respect to ground (AUCg).

In the Rotterdam Study genotyping was performed using Illumina HumanHap 550v3 and Illumina HumanHap 610. The genotyped dataset was restricted to persons who reported that they were from European descent. Ethnic outliers were further excluded by removing samples which showed more than 4SD difference to the study population mean on any of the first 4 dimensions of a multidimensional scaling analysis. We also excluded samples with gender mismatch and excess autosomal heterozygosity as well as duplicates and monozygotic twins (>97% estimated identity-by-descent proportion). Furthermore, second degree cousins or closer relatives were excluded during the GREML analysis by using a GRM cutoff of 0.025 to avoid bias from shared environment. MACH 1.0 software was used to impute to ~30M SNPs based on the 1000 genomes Phase I version 3 reference panel.<sup>23</sup> SNPs included in imputation met the thresholds minor allele frequency >=1%, Hardy-Weinberg equilibrium  $p > 10E-06$ , and a SNP call rate >=98.0%.

## GREML

SNP heritability of the cortisol measurements in the Rotterdam Study were estimated using individual level data with GREML, as implemented in Genome-wide Complex Trait Analysis (GCTA) 1.25.3.<sup>20</sup> GREML quantifies how well the similarity in the genotype between study participants explains the similarity in phenotype. Genetic similarity was established by computing a genetic relatedness matrix (GRM). We used 8,131,668 imputed autosomal SNPs to create the GRM, after filtering for imputation quality ( $R^2 > 0.5$ ) and minor allele frequency (MAF)  $\geq 0.01$ . The GRM was specified as a random effect predicting cortisol levels. To test whether this genetic effect statistically significantly predicts the phenotype, we compared the GRM to a simpler model without the GRM using a likelihood ratio test.

Visual examinations of the total genetic effect and residuals using QQ-plots showed deviations from normality for the saliva measurements. The distribution was normal after square root transformation of hormone levels for saliva cortisol. A constant (+1) was added before transformation to avoid zero values. We report results from analyses on transformed saliva and untransformed plasma levels. Additionally, we performed a power analysis as described by Visscher et al.<sup>24</sup> The plasma cortisol GREML analyses were well powered to detect 16% heritability (power=80% at  $\alpha=0.05$  and  $2E-5$  genetic

relationship). The power to detect SNP heritability was less in the saliva GREML analyses and thus these analyses have less precision.

### **Covariates and Confounders**

We adjusted the phenotype in all analyses for age, sex and four principal components (PC) of ancestry (computed with GCTA). This was achieved by regressing the phenotype on the covariates and using the residuals as outcome in the GREML analysis. The residuals were computed in R 3.2.3.<sup>25</sup> Since plasma cortisol levels were measured in three different Rotterdam Study cohorts, a random intercept on the cohort level was introduced in the regression model of plasma cortisol using the lme4 1.1-10 package.<sup>26</sup>

Additionally, we performed a sensitivity analysis with the plasma data aimed at reducing the environmental variance. This model was adjusted for time and fitted in participants with blood sampling before 11am and no self-reported corticosteroid use (n=4,696). To account for non-linear effects, time-of-day was specified using cubic splines with three degrees of freedom. The residuals, representing time-adjusted plasma levels, were then used in further GREML analyses.

### **CORNET Consortium Plasma and Saliva Cortisol GWAS**

Detailed description of the CORNET GWAS on plasma cortisol can be found in Bolton et al.<sup>19</sup> Briefly, basal morning plasma cortisol was measured in 12,597 participants in 11 western European cohorts. Blood samples were collected between 7am and 11am and analyzed using immunoassays. All participants were at least 17 years old and of European ancestry, were not using glucocorticoids, pregnant, or breast feeding. In total 2945 participants (23%) were included from the Rotterdam Study. However, the measurements were collected in a different study wave than the one used for GREML analyses. HapMap-imputed autosomal SNPs were associated with z-scores of log-transformed plasma cortisol levels in an age, sex and time adjusted additive model. The SNP effects were meta-analyzed with a fixed effect model using inverse-variance weighting. After quality control, the data featured 2,660,191 SNPs with minor allele frequency >2%.

In parallel, an additional GWAS of morning saliva levels was performed. This study is unpublished and therefore is presented in more detail. Morning (at awakening) saliva cortisol was measured in 7,703 participants in 8 cohorts: the British 1958 Birth Cohort-Type 1 Diabetes Genetics Consortium (N=1762); the British 1958 Birth Cohort-Wellcome Trust Case-Control Consortium (N=1052);<sup>27</sup> the Netherlands Study of Depression and Anxiety (N=1220);<sup>28</sup> the Netherlands Twin Register (N=162);<sup>29</sup> the Rotterdam Study I (N=1767); the Rotterdam Study III (N=1119); the Multi-Ethnic Study of Atherosclerosis (N=166);<sup>30</sup> and the Tracking Adolescents' Individual Lives Survey (N=455).<sup>31</sup> Only awakening samples collected before 11 am were included in the analyses. Participants using systemic corticosteroids and pregnant and breast-feeding women were excluded from the analyses. All subjects were at least 16 years old and of European ancestry. Details of the genotyping and imputation are given in Table S2. Genotype quality control was per-

formed in each study separately (HWE P-value  $>10^{-6}$ , MAF  $>0.01$ , SNP-call-rate  $>95\%$ ). A z-score was calculated (cortisol at awakening per SD-score in the cohort) to standardize cortisol measurements across cohorts. A linear regression analysis was performed on z-scores of morning saliva cortisol levels adjusted for sex, age and genetic ancestry (cohort specific) using all imputed SNPs.

The meta-analysis was performed with a fixed-effects inverse variance model using the software METAL.<sup>32</sup> In addition to study-specific pre-imputation quality control, SNPs with a MAF  $<0.05$  and an observed to expected variance ratio (imputation quality) less than 0.3 were excluded at the meta-analysis level. Furthermore, only SNPs with information from 4 or more studies were included, resulting in a final SNP number of 2,156,702 SNPs. Genomic control correction was applied to each study. This GWA morning cortisol saliva meta-analysis has an overlap with the GREML analysis of 1767 participants/measurements (23%) from the Rotterdam Study. QQ and Manhattan plots were created with qqman 0.1.4.<sup>33</sup>

### LD Score Regression

LD Score regression exploits the relationship between SNP-Phenotype association strengths and linkage disequilibrium (LD) patterns.<sup>34</sup> Some SNPs show stronger associations than expected due to chance. Assuming true causal effects, the SNPs which are in higher linkage disequilibrium (LD) with nearby SNPs are expected to have more inflated test statistics, because they are more likely to tag causal variants with stronger effects. This makes it possible to use a LD score of a SNP, defined as the sum of  $r^2$  in a 1cM region, as a predictor of the association strength in a regression. The variance explained by the LD score is equivalent to the SNP heritability estimated by GREML. The advantage of LD score regression is, that it can be conducted with summary data from a GWAS and no individual level information is required. However, this analysis tends to have larger standard errors compared to GREML, which uses individual level data and thus can test SNP heritability effects directly.

The SNP  $h^2$  was estimated using LD score regression 1.0.0<sup>34</sup> in the CORNET GWAS data. Since imputation quality can confound LD score regression results, we restricted the analysis to a list of well-imputed SNPs, as recommended by the software

**Table 1:** Descriptive statistics of the Rotterdam Study cortisol measurements and participant characteristics

<b>Cortisol Phenotype</b>	<b>Median Levels in nmol/l (25%; 75% quantile)</b>	<b>Median Age in years (25%; 75% quantile)</b>	<b>Sex (% female)</b>	<b>Median time of collection in Hr (25%; 75% quantile)</b>
Plasma	345.6 (281.7;418.1)	63.6 (58.2;72.44)	57%	0942 (0900;1030)
Saliva (awakening)	13.15 (8.7;18.8)	74.3 (70.5;78.9)	56%	0730 (0700;0806)
Saliva (AUCg)	7.90 (5.7;10.4)	74.3 (70.5;78.8)	55%	-

authors. After applying default quality control settings (see Table S3), the final SNP number was 1,028,327 for plasma cortisol and 951,308 for saliva cortisol.

## RESULTS

### SNP Heritability

Descriptive statistics of the plasma and saliva cortisol levels can be found in Table 1. SNP heritability estimates were low for all cortisol measurement methods, analytical approaches, and cohorts. See Table 2 for full results.

### Plasma Cortisol

We estimated the SNP heritability of plasma cortisol using individual level data of the Rotterdam Study (n=5,705) with GREML. In this cohort approximately 1% [95%CI: 0-12%] of variance in plasma cortisol could be explained by common autosomal gene variants. Adjusting for time of day and excluding participants with plasma cortisol measurements after 11am or those using corticosteroids did not meaningfully change results.

We further investigated the SNP heritability of plasma cortisol in a larger consortium sample: the CORNET cohorts (ncohorts= 11, nparticipants=12,597). We applied LD score regression to estimate SNP heritability of plasma cortisol across multiple cohorts using

**Table 2:** SNP Heritability estimates of plasma and saliva cortisol measurements.

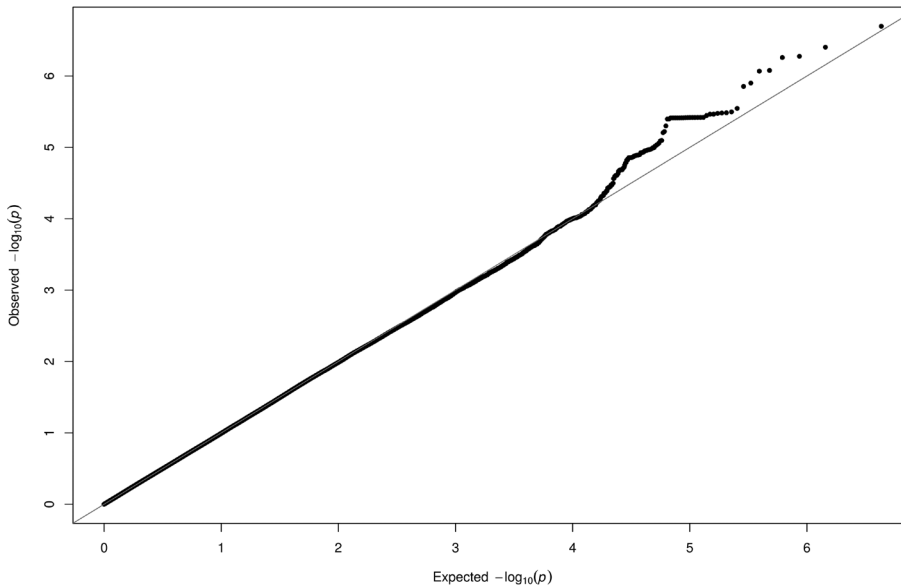
Cortisol Phenotype	Analysis Method	Number of SNPs	n	SNP h <sup>2</sup>	SE	p
<i>Main Analyses:</i>						
Plasma	GREML	8,131,668	5,705	0.006	0.059	0.460
Plasma	LD Score	1,028,327	12,597	0.061	0.035	-
Saliva	GREML	8,131,668	1,717	0.090	0.200	0.329
Saliva (AUCg)	GREML	8,131,668	1,541	0.041	0.210	0.420
Saliva	LD Score	951,308	7,703	-0.083	0.060	-
<i>Sensitivity Analysis:</i>						
Plasma-11am	GREML	8,131,668	4,696	0.000	0.073	0.500

Analyses were adjusted for age, sex and ancestry. Plasma cortisol GREML analyses were further adjusted for cohort effects. Additionally, a sensitivity analysis with adjustment for time-of-day and a subset of participants with measurements before 11am and no reported corticosteroid use is reported (Plasma-11am). Negative heritability values can occur for LD score regression analyses due to sampling variance.

the summary results of a GWAS meta-analysis. The variance explained for this larger sample was also low with 6% [95%CI: 0-13%].

### Saliva Cortisol

In addition to plasma cortisol, we estimated the SNP heritability of two saliva cortisol phenotypes: awakening and diurnal levels. First, we estimated the variance explained of saliva awakening levels in the Rotterdam Study with GREML (n=1,717). The heritability in this sample was 9% [95%CI: 0-48%]. Repeating the analysis in the larger CORNET sample (ncohorts= 8, nparticipants=7,703) using LD score regression on GWAS meta-analysis summary statistics showed a negative heritability estimate (-0.0833). Phenotypes with low heritability can be estimated as negative due to sampling variance, which suggests population heritability close to 0 and an upper 95% confidence interval of 3%. Finally, we estimated the SNP heritability of diurnal cortisol levels (AUCg). These were only



**Figure 1:** Quantile-quantile plot of observed  $-\log_{10} p$  values vs expected  $-\log_{10} p$  values assuming chance findings. Diagonal line indicates a  $p$  value distribution compatible with chance finding. Upward deviations indicate  $p$  values more significant than expected.

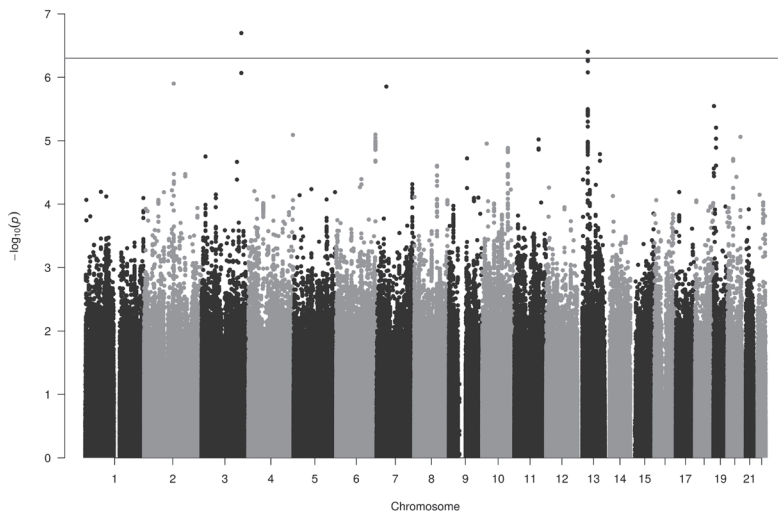
available in the Rotterdam Study ( $n=1,541$ ). In this sample the heritability was estimated at 4% [95%CI: 0-45%].

### Morning Plasma and Saliva Cortisol GWAS

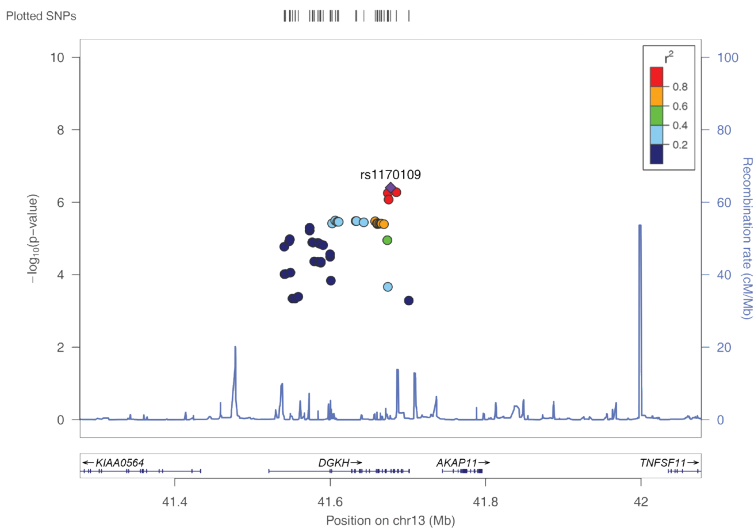
The CORNET GWAS meta-analysis of plasma cortisol, which was previously published,<sup>19</sup> identified 4 SNPs in the SERPINA6/SERPINA1 locus, namely rs12589136, rs2749527, rs2749529 and rs11621961.

However, no SNP reached genome-wide significance ( $p < 5 \times 10^{-8}$ ) in the GWAS for awakening saliva cortisol. Figure 2 displays a Manhattan plot. Two loci showed suggestive associations ( $p < 5 \times 10^{-7}$ ). The T allele of rs1170109 (chr13:42779694) was associated with a 0.12 SD increase in cortisol levels ( $SE=0.02$ ,  $p=3.95 \times 10^{-7}$ ,  $MAF=12\%$ ,  $n=7,690$ ) with a homogeneous effect across the cohorts ( $I^2=0\%$ ). Several SNPs from the same locus, close to the gene DGKH, showed suggestive effects as well (see Figure 3 for a LocusZoom plot<sup>35</sup>). The locus was not associated with plasma cortisol ( $\beta=0.03$ ,  $SE=0.02$ ,  $p=0.17$ ,  $I^2=0\%$ ,  $n=12,592$ ). In the second locus, the A allele of rs6768297 (chr3:168334386) was associated with 0.34 standard deviations (SD) lower cortisol levels ( $SE=0.06$ ,  $p=2.01 \times 10^{-$





**Figure 2:** Manhattan plot of  $-\log_{10} p$  values vs SNP position. SNPs above the horizontal line indicate suggestive findings ( $p < 5 \times 10^{-7}$ ).



**Figure 3:** Regional plot around lead SNP rs1170109.  $-\log_{10} p$  values of rs1170109 and other top1000 SNPs in the region are displayed color coded for strength of correlation.

7). Furthermore, the SNP showed a nominally significant ( $\alpha=0.05$ ) association with plasma cortisol in the same direction ( $\beta=-0.08$ ,  $SE=0.03$ ,  $p=0.01$ ,  $I^2=0\%$ ,  $n=11,441$ ). Rs6768297 had a low MAF (6%), high effect heterogeneity ( $I^2=85.5\%$ ) and information was only available in 40% of the sample ( $n=3054$ ). None of the four SNPs associated with plasma cortisol were associated with saliva cortisol (all  $p<0.56$ ).

The LD score intercept was 1.0031 ( $SE=0.0066$ ) and 1.0085 ( $SE=0.0073$ ) for the plasma and saliva GWAS, respectively, suggesting no inflation due to population stratification. The QQ plots also showed no problematic inflation (see Figure 1 for saliva).

## DISCUSSION

The low heritability of plasma cortisol in two large samples estimated by two different approaches strongly suggests that plasma cortisol is not substantially affected by the additive effects of autosomal SNPs. The same conclusion can be drawn for morning saliva cortisol, which was also estimated by two analytical approaches, and to a lesser extent for diurnal cortisol.

No SNP reached genome-wide significance in a GWAS of morning saliva cortisol levels, which is expected for traits with low SNP heritability analyzed in relatively small samples. Two loci showed suggestive associations. Interestingly, one top SNP rs6768297 lies within the EGFEM1P gene, which has a high and specific expression in the pituitary according to RNA expression data (1.5 reads per kilobase per million).<sup>36,37</sup> Furthermore, the SNP showed a nominally significant association with plasma cortisol in the same direction as saliva cortisol.

However, the lack of genome-wide significance, low sample size, low MAF and high effect heterogeneity also cast doubt as to whether the rs6768297 association with cortisol would replicate in a completely independent sample. The SERPINA6/SERPINA1 locus identified in the plasma cortisol GWAS19 appears to be specific to plasma cortisol levels.

The results are consistent with phenotypic studies indicating that only a small proportion of cortisol variance shows a stable trait-like pattern. In three different studies Ross et al.<sup>38</sup> found that 44.4%-75.5% of total day-time cortisol output variance was under day-to-day fluctuations. Studying children through ages 9-15, Shirtcliff et al.<sup>39</sup> found that situation-specific environmental influences can explain 52% of cortisol variance (excluding circadian rhythm). The authors conclude that only 13% of the cortisol variance at a given time shows trait-like stability over the years, which coincides with the upper confidence intervals found for the heritability of acute plasma levels. These studies highlight the fact that cortisol secretion and metabolism is a highly dynamic process

adapting to not only short-term, but also long-term situational contexts, which results in considerable “noise” in genetic studies.

This notion is supported by the low heritability of the diurnal cortisol measurements. Reducing the within-day variation appears to be insufficient to reduce the contextual noise. This conclusion is further supported by the small effect adjusting for time-of-day had on the plasma cortisol estimates and the low heritability of awakening saliva cortisol. The latter has a precise circadian definition, though sampling can be difficult to time in a home environment. Furthermore, after excluding participants with plasma cortisol measurements after 11am and corticosteroid use, heritability estimates remained under 1%.

Interestingly, long-term associations between single cortisol measures in adulthood and psychosocial problems and adversities in childhood have been found.<sup>40,41</sup> The variability might thus reflect environmental exposures, but for genetic studies more long-term profiles of cortisol may be needed. These can be measured using hair samples, which might represent more trait-like effects with less environmental influence.<sup>42,43</sup> However, long-term environmental contexts spanning months or years also contribute to the cortisol variance and it is unclear yet to what extent 3 to 6 month measurements shall reduce environmental noise.

Therefore there may not be a single simplistic genomic heritability of cortisol levels. It is tempting to speculate that the heritability of other cortisol phenotypes is higher. Indeed the reliability of, for example, the total daily cortisol values (AUCg) is higher than single morning samples,<sup>13,14</sup> but it represents a distinct feature of the cortisol secretion pattern. The cortisol awakening response or diurnal slopes are two other examples of characterizing diurnal changes. These may show a different balance of genetic and environmental influences than total daily values or hair cortisol. The awakening response or diurnal slopes may show higher heritability than the tested phenotypes, though, it should be noted that they show less stability than total daily output.<sup>38</sup> Another potentially interesting phenotype is cortisol reactivity to various stressors. Here again the heritability may be different and may even change depending on the stressor. Unfortunately, sample sizes for stress reactivity will likely be smaller. Future research is required to determine the SNP heritability of these alternative phenotypes and characterize potential differences between them, although this may be a challenging research field.

The very low diurnal cortisol heritability is in line with a twin-study reporting no genetic effects for day-time profiles.<sup>18</sup> The same study found a non-significant heritability of 26% for awakening cortisol, which is compatible with the non-significant point estimate of 9% SNP heritability in the GREML analysis. Further, the observed 0% to 6% SNP heritability for (mostly morning) plasma and saliva levels (LD score regression) are similar to the 0% and 14% twin heritabilities reported for saliva and plasma morning

levels.<sup>15,18</sup> However, they show a substantial difference to twin studies finding a 45% heritability of acute plasma levels.<sup>16,17</sup>

SNP heritability is expected to be lower than twin heritability, since this estimate does not include the effects of rare, structural and X-linked variants, which are captured in twin studies. Gene-gene and gene-environment interactions can also substantially increase standard twin heritability estimates.<sup>44</sup> Alternatively, 45% twin heritability of acute cortisol measurements might be an overestimation, which would be consistent with the fact that the twin studies are highly inconsistent.

The LD score regression and GREML analysis of plasma cortisol in the CORNET and Rotterdam Study samples had good power to detect modest heritability. The negative findings in addition to the convergent evidence from the smaller saliva cortisol samples suggest that acute cortisol measures have low SNP heritability. However, the evidence is less clear for day-time profiles. These were only available in a small sample and have very wide confidence intervals, thus firm conclusions cannot be made. Another limitation is that the CORNET and Rotterdam Study data have an overlap in participants of approximately 20%. The samples were thus not completely independent. However, considering that the majority of the observations did not overlap and the measurements were taken at different times and assessed in different laboratories, the data nevertheless support robustness of the largely negative results.

The findings suggest that common autosomal SNPs are poor predictors of acute cortisol levels. However, predictive power is not equal to importance. Crucial cortisol regulating loci are highly conserved: mammals and fish have a similar stress physiology. Among others, corticotrophin-releasing hormone genes are orthogonal with substantial overlap in amino acid identity.<sup>45</sup> This highlights the importance of cortisol related genes, but also suggests that natural selection restricts the amount of variation and in turn effect sizes and predictive power. This may suggest, that if SNPs are identified despite the low SNP heritability, such as SNPs of the SERPINA6/SERPINA1 locus in the plasma cortisol GWAS, they are all the more important.

Unfortunately, it follows from the presented results, that detecting these SNPs will be difficult. Since most SNPs are expected to have a relatively low predictive contribution compared to the environment and stochastic factors, very large sample sizes are probably required to discover further loci. Given the apparent importance of cortisol genetics, GWAS seems nevertheless a worthwhile endeavor to uncover further cortisol related biological pathways.

## REFERENCES

1. West DWD, Phillips SM. Associations of exercise-induced hormone profiles and gains in strength and hypertrophy in a large cohort after weight training. *Eur J Appl Physiol.* 2012;112(7):2693-2702. doi:10.1007/s00421-011-2246-z
2. Barton R, Stoner H, Watson S. Relationships among plasma cortisol, adrenocorticotrophin, and severity of injury in recently injured patients. *J Trauma.* 1987;27(4):384-392. [http://journals.lww.com/jtrauma/Abstract/1987/04000/Relationships\\_among\\_Plasma\\_Cortisol,.7.aspx](http://journals.lww.com/jtrauma/Abstract/1987/04000/Relationships_among_Plasma_Cortisol,.7.aspx). Accessed January 29, 2016.
3. Kudielka BM, Hellhammer DH, Wüst S. Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. *Psychoneuroendocrinology.* 2009;34(1):2-18. doi:10.1016/j.psyneuen.2008.10.004
4. Adam EK. Transactions among adolescent trait and state emotion and diurnal and momentary cortisol activity in naturalistic settings. *Psychoneuroendocrinology.* 2006;31(5):664-679. doi:10.1016/j.psyneuen.2006.01.010
5. Young E, Abelson J, Lightman S. Cortisol pulsatility and its role in stress regulation and health. *Front Neuroendocrinol.* 2004;25(2):69-76. doi:10.1016/j.yfrne.2004.07.001
6. Gozansky WS, Lynn JS, Laudenslager ML, Kohrt WM. Salivary cortisol determined by enzyme immunoassay is preferable to serum total cortisol for assessment of dynamic hypothalamic-pituitary-adrenal axis activity. *Clin Endocrinol (Oxf).* 2005;63(3):336-341. doi:10.1111/j.1365-2265.2005.02349.x
7. Restituto P, Galofré JC, Gil MJ, et al. Advantage of salivary cortisol measurements in the diagnosis of glucocorticoid related disorders. *Clin Biochem.* 2008;41(9):688-692. doi:10.1016/j.clinbiochem.2008.01.015
8. Ruttle PL, Javaras KN, Klein MH, Armstrong JM, Burk LR, Essex MJ. Concurrent and longitudinal associations between diurnal cortisol and body mass index across adolescence. *J Adolesc Heal.* 2013;52(6):731-737. doi:10.1016/j.jadohealth.2012.11.013
9. Walker BR. Glucocorticoids and cardiovascular disease. *Eur J Endocrinol.* 2007;157(5):545-559. doi:10.1530/EJE-07-0455
10. Yehuda R, Seckl J. Minireview: Stress-Related Psychiatric Disorders with Low Cortisol Levels: A Metabolic Hypothesis. *Endocrinology.* 2011;152(12):4496-4503. doi:10.1210/en.2011-1218
11. Girshkin L, Matheson SL, Shepherd AM, Green MJ. Morning cortisol levels in schizophrenia and bipolar disorder: A meta-analysis. *Psychoneuroendocrinology.* 2014;49(1):187-206. doi:10.1016/j.psyneuen.2014.07.013
12. Fischer S, Strawbridge R, Herane Vives A, Cleare AJ. Cortisol as a predictor of psychological therapy response in depressive disorders: systematic review and meta-analysis. *Br J Psychiatry.* 2016. doi:10.1192/bjp.bp.115.180653
13. Golden SH, Wand GS, Malhotra S, Kamel I, Horton K. Reliability of hypothalamic-pituitary-adrenal axis assessment methods for use in population-based studies. *Eur J Epidemiol.* 2011;26(7):511-525. doi:10.1007/s10654-011-9585-2
14. Elder GJ, Ellis JG, Barclay NL, Wetherell MA. Assessing the daily stability of the cortisol awakening response in a controlled environment. *BMC Psychol.* 2016;4(1):3. doi:10.1186/s40359-016-0107-6
15. Froehlich JC, Zink RW, Li TK, Christian JC. Analysis of heritability of hormonal responses to alcohol in twins: beta-endorphin as a potential biomarker of genetic risk for alcoholism. *Alcohol*

- Clin Exp Res. 2000;24(3):265-277. <http://www.ncbi.nlm.nih.gov/pubmed/10776662>.
16. Inglis GC, Ingram MC, Holloway CD, et al. Familial pattern of corticosteroids and their metabolism in adult human subjects - The Scottish adult twin study. *J Clin Endocrinol Metab.* 1999;84(11):4132-4137.
  17. Meikle a W, Stringham JD, Woodward MG, Bishop DT. Heritability of variation of plasma cortisol levels. *Metabolism.* 1988;37(6):514-517. <http://www.ncbi.nlm.nih.gov/pubmed/2967419>.
  18. Wüst S, Federenko I, Hellhammer DH, Kirschbaum C. Genetic factors, perceived chronic stress, and the free cortisol response to awakening. *Psychoneuroendocrinology.* 2000;25(7):707-720. doi:10.1016/S0306-4530(00)00021-4
  19. Bolton JL, Hayward C, Direk N, et al. Genome wide association identifies common variants at the SERPINA6/SERPINA1 locus influencing plasma cortisol and corticosteroid binding globulin. *PLoS Genet.* 2014;10(7):e1004474. doi:10.1371/journal.pgen.1004474
  20. Yang J, Lee S, Goddard M, Visscher P. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet.* 2011;88(1):76-82. doi:DOI 10.1016/j.ajhg.2010.11.011
  21. Hofman A, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol.* 2015;30(8):661-708. doi:10.1007/s10654-015-0082-x
  22. Dekker MJHJ, Koper JW, Van Aken MO, et al. Salivary cortisol is related to atherosclerosis of carotid arteries. *J Clin Endocrinol Metab.* 2008;93(10):3741-3747. doi:10.1210/jc.2008-0496
  23. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74. doi:10.1038/nature15393
  24. Visscher PM, Hemani G, Vinkhuyzen AAE, et al. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* 2014;10(4):e1004269. doi:10.1371/journal.pgen.1004269
  25. R Core Team. R: A Language and Environment for Statistical Computing. 2015. <http://www.r-project.org>.
  26. Bates D, Mächler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models using lme4. *J Stat Softw.* 2014;1:1-51. <http://arxiv.org/abs/1406.5823>[http://listengine.tuxfamily.org/lists.tuxfamily.org/eigen/2011/06/pdfKU\\_S0z6LjT.pdf](http://listengine.tuxfamily.org/lists.tuxfamily.org/eigen/2011/06/pdfKU_S0z6LjT.pdf)<https://github.com/lme4/lme4/blob/master/inst/doc/lmer.pdf><https://publication/uuid/8D147730-46C2-4182-9F9A-505117201D22>.
  27. Power C, Li L, Hertzman C. Associations of early growth and adult adiposity with patterns of salivary cortisol in adulthood. *J Clin Endocrinol Metab.* 2006;91(11):4264-4270. doi:10.1210/jc.2006-0625
  28. Penninx BWJH, Beekman ATF, Smit JH, et al. The Netherlands Study of Depression and Anxiety (NESDA): Rationale, objectives and methods. *Int J Methods Psychiatr Res.* 2008;17(3):121-140. doi:10.1002/mp.256
  29. Boomsma DI, Geus EJC De, Vink JM, et al. Netherlands Twin Register : From Twins to Twin Families. *Twin Res Hum Genet.* 2006;9(6):849-857. doi:10.1375/twin.9.6.849
  30. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and design. *Am J Epidemiol.* 2002;156(9):871-881. doi:10.1093/aje/kwf113
  31. Huisman M, Oldehinkel AJ, de Winter A, et al. Cohort profile: The Dutch TRacking Adolescents Individual Lives Survey; TRAILS. *Int J Epidemiol.* 2008;37(6):1227-1235. doi:10.1093/ije/dym273
  32. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340

33. Turner SD. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *bioRxiv*. 2014:5165.
34. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;advance on(3):291-295. doi:10.1038/ng.3211
35. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*. 2011;27(13):2336-2337. doi:10.1093/bioinformatics/btq419
36. GTEx-Portal. GTEx Portal. <https://www.gtexportal.org/home/>. Published 2017. Accessed May 11, 2017.
37. Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585.
38. Ross KM, Murphy MLM, Adam EK, Chen E, Miller GE. How stable are diurnal cortisol activity indices in healthy individuals? Evidence from three multi-wave studies. *Psychoneuroendocrinology*. 2014;39:184-193. doi:10.1016/j.psyneuen.2013.09.016
39. Shirtcliff E a., Allison AL, Armstrong JM, Slattery MJ, Kalin NH, Essex MJ. Longitudinal stability and developmental properties of salivary cortisol levels and circadian rhythms from childhood to adolescence. *Dev Psychobiol*. 2012;54(5):493-502. doi:10.1002/dev.20607
40. Power C, Li L, Atherton K, Hertzman C. Psychological health throughout life and adult cortisol patterns at age 45y. *Psychoneuroendocrinology*. 2011;36(1):87-97. doi:10.1016/j.psyneuen.2010.06.010
41. Power C, Thomas C, Li L, Hertzman C. Childhood psychosocial adversity and adult cortisol patterns. *Br J Psychiatry*. 2012;201(3):199-206. doi:10.1192/bjp.bp.111.096032
42. Rippe RCA, Noppe G, Windhorst DA, et al. Splitting hair for cortisol? Associations of socio-economic status, ethnicity, hair color, gender and other child characteristics with hair cortisol and cortisone. *Psychoneuroendocrinology*. 2015;66:56-64. doi:10.1016/j.psyneuen.2015.12.016
43. Noppe G, de Rijke YB, Dorst K, van den Akker ELT, van Rossum EFC. LC-MS/MS based method for long-term steroid profiling in human scalp hair. *Clin Endocrinol (Oxf)*. 2015;83(2):162-166. doi:10.1111/cen.12781
44. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A*. 2012;109(4):1193-1198. doi:10.1073/pnas.1119675109
45. Huising MO, Metz JR, van Schooten C, et al. Structural characterisation of a cyprinid (*Cyprinus carpio* L.) CRH, CRH-BP and CRH-R1, and the role of these proteins in the acute stress response. *J Mol Endocrinol*. 2004;32(3):627-648. doi:10.1677/jme.0.0320627

## SUPPLEMENTAL TABLES

**Table S1:** Sample characteristics for CORNET cohorts participating in plasma cortisol GWAS (from Bolton et al. (2014)) and saliva cortisol GWAS.

Study	n	Men (%)	Age in years		Cortisol (nmol/l)		Sampling clock time
			M (SD)	Range	M (SD)	Range	Hr
<i>Plasma Cortisol</i>							
ORCADES	886	45	53.5(15.7)	17-97	765 (315)	Nov-41	0830-1030
CROATIA-Korcula	898	36	56.2 (13.9)	18-98	698 (207)	59-815	0800-0900
CROATIA-Split	496	43	45.0 (14.7)	18-85	979 (404)	94-2831	0730-0900
CROATIA-Vis	892	44	56.4 (15.5)	18-93	622 (230)	64-1820	0730-0900
RS1	2945	45	71.9 (7.0)	61-105	305 (94)	5-679	0800-1100
HBCS1934-44	451	36	60.61 (2.80)	56-67	393 (120)	125-990	0750-1055
NFBC1966	1192	0	31(0)	n/a	380 (160)	40-2370	0800-1100
ALSPAC	1567	50	15.43 (0.26)	14-17	486 (174)	58-1683	0800-1057
InChianti	1210	45	68.3 (15.6)	21-102	375 (135)	19-1291	Before 0900
PIVUS	919	50	70.2 (0.17)	69-72	386 (125)	31-930	0800-1000
PREVEND	1151	51	49.4 (13.0)	28-75	442 (201)	20-1734	0800-1100
<i>Saliva Cortisol</i>							
B58C-T1DGC	1762	49	45.3 (0.34)	45-46	21.1 (10.4)	3.7-61.6	0500-1100
B58C-WTCCC	1052	49	44.9 (0.35)	45- 46	21.0 (9.8)	3.7-60.8	0500-1100
NESDA	1220	33	43.5 (12.5)	18-65	17.3 (7.98)	2.0-66.6	0250-1100
NTR	162	29	29.4 (12.4)	14-75	15.8 (7.4)	2.0-46.0	0800-1100
RS1	1767	44	75.0 (5.8)	65-98	14.6 (8.5)	0.0-60.1	0248-1100
RS3	1119	43	55.3 (4.8)	46-69	14.8 (8.7)	0.7-72.0	0500-1100
MESA	166	50	67.5 (10.3)	49-90	16.2 (9.3)	2.6-89.8	Before 1100
TRAILS	455	47	16.1 (0.6)	14-17	10.9 (6.5)	1.0-112.0	0300-1100

Reference: Bolton, J.L., Hayward, C., Direk, N., Lewis, J.G., Hammond, G.L., Hill, L.A., Anderson, A., Huffman, J., Wilson, J.F., Campbell, H., Rudan, I., Wright, A., Hastie, N., Wild, S.H., Velders, F.P., Hofman, A., Uitterlinden, A.G., Lahti, J., Räikkönen, K., Kajantie, E., Widen, E., Palotie, A., Eriksson, J.G., Kaakinen, M., Järvelin, M.-R., Timpson, N.J., Davey Smith, G., Ring, S.M., Evans, D.M., St Pourcain, B., Tanaka, T., Milaneschi, Y., Bandinelli, S., Ferrucci, L., van der Harst, P., Rosmalen, J.G.M., Bakker, S.J.L., Verweij, N., Dullaart, R.P.F., Mahajan, A., Lindgren, C.M., Morris, A., Lind, L., Ingelsson, E., Anderson, L.N., Pennell, C.E., Lye, S.J., Mathews, S.G., Eriksson, J., Mellstrom, D., Ohlsson, C., Price, J.F., Strachan, M.W.J., Reynolds, R.M., Tiemeier, H., Walker, B.R., 2014. Genome wide association identifies common variants at the SERPINA6/SERPINA1 locus influencing plasma cortisol and corticosteroid binding globulin. *PLoS Genet.* 10, e1004474. doi:10.1371/journal.pgen.1004474



**Table S2:** Morning Saliva GWAS: genotyping methods per cohort

<b>Study</b>	<b>N</b>	<b>Genotyping Platform</b>	<b>Calling Algorithm</b>	<b>Reference Panel</b>
B58C-T1DGC	1762	III 550K	Illuminus	HapMap release 21 CEU (build 35)
B58C-WTCCC	1052	Aff 500K	Chiamo	HapMap release 21 CEU (build 35)
NESDA	1220	Perlegen 600K	Birdseed	HapMap release 22 CEU (build 36)
NTR	162	Perlegen 600K	Beadstudio	HapMap release 22 CEU (build 36)
RS1	1767	Illumina Human Map 550K	Beadstudio	HapMap release 22 CEU (build 36)
RS3	1119	Illumina Human Map 610K	Beadstudio	HapMap release 22 CEU (build 36)
MESA	166	Aff 6.0	Birdseed	HapMap release 24 CEU (build 36)
TRAILS	455	Illumina BeadStation 500	Beadstudio	HapMap release 22 CEU (build 36)

V

**Table S3:** Quality Control (QC) applied to SNPs before LD score regression

<b>QC step</b>	<b>Plasma</b>	<b>Saliva</b>
Original SNP number	2,660,191	2,472,180
Overlap well-imputed SNP List	- 1,517,683	- 1,398,157
Missing Values	0	0
MAF <= 0.01	0	0
Out-of-bounds p-values	0	0
Strand-ambiguity	-7	-4
Duplicated SNPs	0	0
Low sample size	-113,994	-122,663
Mismatched alleles	-29	-23
Missing LD scores	-151	-24
Final SNP number	1,028,327	951,308







Chapter VI

**General Discussion**

## GENERAL DISCUSSION

In this thesis I investigated the etiology and biological correlates of general psychopathology, DNA methylation in relation to ADHD symptoms and the genetics and biases of cortisol measurements. The studies confirmed some of our hypothesis, rejected others and produced some unexpected results. In the following the findings of each chapter will be summarized and interpreted. I will then discuss methodological considerations, followed by clinical implications and a thesis summary. Rather than bore the readers and repeat the discussion provided in previous chapters, I selected several findings for more in depth considerations.

### Chapter III

In chapter III.A we investigated the association of parental age at delivery with child internalizing and externalizing symptoms in school-age. This study rejects the hypothesis that parental age is linked to internalizing problems. The conclusions was highly consistent across all four cohorts participating in the study. It appears that parental age either does not have adverse impacts or that the biological aspects of parental are compensated by other associated factors. A natural candidate for compensation would be better socioeconomic status (SES) with higher age, though adjustment for indicators of SES did not change conclusions. Surprisingly, higher parental age had beneficial associations with externalizing problems. Again, adjusting for SES did not change conclusions. Since the associations were only present for externalizing symptoms, one may conclude that the parental age effects are exclusive to externalizing disorders. However, without testing general psychopathology factor models, this conclusion may be premature as chapter III.5 teaches us. Using an internalizing/externalizing model without including the general psychopathology factor did not suggest any associations in that chapter, while associations with general and specific externalizing factors were detected using a general psychopathology factor model. Thus, parental age may also be associated with internalizing symptoms indirectly via general psychopathology, but future research is needed to investigate this possibility.

Chapters III.B and III.C and III.E introduced general psychopathology factors based on multiple informants in school-aged children. These models were first evaluated based on fit statistics, i.e. measures of how well they explain the observed correlation between the symptom scores. Then the psychopathology factors associated were associated with external predictors, childhood correlates and adolescence/adulthood outcomes. Specifically we tested associations with single nucleotide polymorphism, white matter integrity at age 10, school achievement test results, and with criminal behavior in adulthood, problem drinking, wellbeing and psychiatric diagnoses.

In general, we found that child psychiatric symptoms can be modeled as a consequence of both general and specific factors, as models including both general psycho-

pathology factor and specific factors (internalizing, externalizing/attention) simultaneously better explain the correlation between symptom scores compared to models of general or domain specific factors only. This finding was consistent among three cohorts of the DREAM BIG consortium. Importantly, we demonstrated that these latent factors have meaningful biological correlates and predictive validity. The general psychopathology was associated with single nucleotide polymorphisms (which explained 36% of its variance) and with less concurrent white matter integrity. General psychopathology had also predictive power, as children with higher levels of general psychopathology in childhood had a higher risk of depression in adulthood, lower well-being, more problem drinking and lower grades on high school completion exams.

At the beginning of this PhD project the bifactor models of psychopathology had been already tested in several cohorts, with the consistent result that the bifactor model has better fit than traditional models of psychopathology with correlated domain specific latent variables.<sup>1-3</sup> In other words, adding a general psychopathology factor to a traditional internalizing/externalizing model, better explained co-occurrence of psychiatric symptoms than only accounting for the shared variance of internalizing and externalizing factors by a correlation. However, several concerns questioned the validity of these studies. One concern was that shared method variance inflated fit statistics. Early papers, e.g.<sup>1,2,4</sup>, fitted bifactor models to observations by a single informant. It is therefore likely that consistent ratings across all symptoms was not only due to real co-occurrence of symptoms, but also tendencies of the informant to rate higher or lower regardless of domain. Second, Bonifay & Cai<sup>5</sup> argue that bifactor models tend to have a bias towards better fit statistics. Fit indices do not appropriately account for the increase in functional complexity, the increased flexibility of bifactor models to fit any data independent of the number of parameters. Third, some researchers propose that bifactor models actually reflect network models.<sup>6</sup> In network models different symptoms can influence each other and do not necessarily originate from common factors.

In regards to the first concern about shared method variance, the risks should be reduced in this thesis, because all general psychopathology factors models were based on multiple informants and assessment time points (see methodological considerations for more further discussion). This protects against shared method variance in several ways. By having repeated measurements, in some cases with different instruments, acute situational biases affecting all symptom rating are averaged out, under the assumption that there is no systematic bias across time. An example would be a parent who is upset and frustrated about the child on the day of the assessment. This could lead to the parent rating all items indiscriminately higher without paying much attention to the specific questions asked. Ratings from multiple informants can also reduce the effect of time invariant biases, such as an overly concerned parent always rating any symptom higher independent of the behavior of the child. Higher ratings across multiple symptoms would only be taken into account if they are consistent among multiple informants. While our multi-informant approach is arguably an improvement over single-informant

designs, it also has some limitations. On the one hand, it does not protect against biases inherent to any informant, such as being more sensitive to noticing any kind of symptom if a specific symptom has previously occurred. On the other hand, the multi-informant approach may also be too stringent and adjust for unique insights each informant has, as only consistent ratings are taken into account.

As for the second point, that conventional fit criteria are inadequate to account for bifactor models' functional complexity: whatever fit criteria is used, the choice of the best model cannot be judged based on statistical measures of fit alone. It is important to validate any construct with external variables, preferably with variables which can be objectively measured. In this thesis, we found that general psychopathology is associated with common SNPs, as well as with global white matter integrity. The usefulness of the general psychopathology model is especially apparent in the latter case, as white matter integrity did not associate with the traditional psychopathology factors internalizing, externalizing and attention. Thus one could have come to the erroneous conclusion that global white matter is not associated with psychopathology when analyzing psychopathology domains separately, while using a general psychopathology factor model the conclusion is that all these domains are in fact indirectly associated through general effects on general psychopathology.

The last concern, that the actual structure of psychopathology is a network and psychopathology is not the direct result of a general psychopathology factor is harder to refute. If psychopathology is best represented by network mechanisms, a bifactor model would still fit well. As an example, it is conceivable that emotional reactivity makes one more likely to react with aggression to a stressor. However, aggression may in turn lead to social repercussions, which may affect depressive symptoms, leading yet again to higher emotional reactivity. A bifactor model may suggest a general psychopathology factor underlying all three symptoms, even though in the example aggression and depression are not caused independently but result from one preceding symptom. To distinguish which model, common factor or network, better describes psychiatric disorders is highly challenging. However, even if psychopathology follows a network structure, bifactor models may still be highly useful, if a simplification. The relationship between latent variable models should perhaps be not understood as necessarily direct and independent, but perhaps should be interpreted more flexibly, as there might be mediation and feedback mechanisms between the items. While for a complete picture network modeling may be necessary, the demands for data with high temporal resolution and complexities in analysis, may make latent factor models a more practical choice for many study designs. For instance, latent variable models are well suited for genetic studies, as feedback on genetic variants from psychopathology is impossible. Furthermore, not all research questions necessarily require insights whether an exposure impacts symptoms directly or via other symptoms.

In chapter III.3, we found that single nucleotide polymorphisms explained 36% of the general psychopathology factor variance (SNP heritability). To identify the specific



loci involved, we performed a genome-wide association study of a total psychiatric sum score, as a proxy of a general psychopathology factor (chapter III.4). We did not identify any specific SNPs associations in the genome-wide association study of total psychiatric problems. This does not necessarily contradict the hypothesis that SNPs are associated with general psychopathology, as the SNP heritability was 8.4%, but it does demonstrate that even sample sizes of almost 30,000 participants are not sufficient to detect SNP specific effects. Using gene-based tests, we did, however, identify the myotonic dystrophy (DM1) gene cluster as associated with total child psychiatric problems. While this locus is known for a rare mutation leading to a correspondingly rare neuromuscular disorder, common variants in this variant appear to have consequences. While common complex genetic traits and rare genetic disorders are often viewed as separate research lines, this example highlights, that both fields can potentially inform each other. Genetic correlation analyses confirmed that common psychiatric disorders co-occur in part due to shared genetic effects. Curiously, that less common psychiatric disorders such as schizophrenia or bipolar disorder co-occur due to shared genetic risk with common disorders was not confirmed. As these disorders are very rare in childhood, they were not represented well in any of the general psychopathology factor models. However, other studies in adulthood did include psychotic and manic symptoms,<sup>2</sup> and found that disorders, such as schizophrenia load on general psychopathology as well. It therefore appears on first glance that the genetic correlations do not follow the phenotypic co-occurrence. In other words, an individual with e.g. depression is more likely to also suffer from schizophrenia at some point in their life compared to an individual with no psychopathology. On the other hand, genetic risk for common child psychiatric symptoms appears to not predict a higher chance for psychotic symptoms. Thus, the co-occurrence of common and less common symptoms such as depression and schizophrenia must be due to a common environmental cause. However, a perhaps more likely explanation for the discrepancy is that a total psychiatric sum score in childhood is not the best measure of truly general psychopathology and needs to include measurements of thought disorder symptoms at later ages. This observation makes the high genetic correlations with other traits, such as smoking behavior, body fat and intelligence even more remarkable, as they were not included in the computation of the total psychiatric sum score. This implies, that a genetic risk for total child psychiatric problems is also a risk for many other medical problems, as the genetic risk for child psychopathology appears to also affect health related behaviors, such as smoking or overeating, which then could result in poorer mental and physical health.

## Chapter IV

Next to new insights into general psychopathology, we investigated the epigenetics of ADHD examining methylation profiles at birth and school age using an epigenome-wide association study. We identified 9 genome-wide significant probes, which were differentially methylated at birth but not at school-age. Two of the probes lie in the genes

*ERC2* and *CREB5*. Both are expressed in the brain and related to neural functioning or development. *ERC2* regulates neurotransmitter release and *CREB5* has important neurite growth functions. In addition, *CREB5* has been previously associated with ADHD. Due to the role in neural functioning, both probes are interesting candidates for etiologically relevant epigenetic regulations of ADHD symptoms. The association of these probes did not persist into school-age, in fact, no probe reached genome-wide significance at school-age. As discussed in Chapter IV, the overall signal of school-age methylation was lower than birth methylation, despite similar sample size. This observation was reported before for ADHD,<sup>7</sup> thus it seems to be a robust finding for this disorder. However, it is much less clear whether cord blood methylation has also a stronger signal for other symptoms. ADHD is a neurodevelopmental disorder with onset in early childhood. It may therefore be, that ADHD symptoms are more sensitive to early prenatal exposures, which are reflected by cord blood methylation or directly caused by perinatal DNA methylation profiles, than other symptoms. However, while other disorders, such as depression or thought disorders commonly have their incidence in adolescence or later, this does not exclude them from being influenced by early exposures. It could be that methylation profiles at birth are more important than at other timepoints for psychopathology in general. If this were the case, what are the consequences for the design of epigenetic studies and the choice of assessment periods?

On the one hand, cord blood methylation appears to have a better price to cost ratio, as under the above assumptions we can expect to find more differentially methylated regions given the same number of arrays. On the other hand, one of the promises of DNA methylation research is that changes in methylation could illuminate the biological pathways linking postnatal environment to psychopathology. This latter issue can only be addressed by repeated measurements throughout the lifespan. As DNA methylation arrays remain more costly than genetic arrays, it may for now be worthwhile to concentrate on birth methylations. Yet the case could be made that investigating biological mediation of environmental risks is of particular importance, and that this challenge should be taken up even if the bar is higher.

## Chapter V

In chapter V.1 we investigated, whether polygenic scores of hair color could predict cortisol levels in hair to assess whether hair color may bias cortisol measurement using hair samples. Indeed, we observed that darker hair is associated with higher cortisol levels independent of genetic ancestry. In chapter V.2 we investigated the SNP heritability of acute saliva and cortisol measures. Neither morning levels nor total day output were related to common single nucleotide polymorphisms.

It was surprising how difficult it is to study the genetics of cortisol. Psychiatry and psychology are often criticized as being a highly subjective field with questionable unverifiable constructs. Within the field the notion exists that objectively measured endophenotypes can facilitate genetic research, as endophenotypes are “closer” to ge-

netics and thus lead to more scientific successes.<sup>8</sup> However, in our studies and others, subjective measures of psychopathology in most cases show consistently higher SNP heritability than 6%, i.e. the SNP heritability of acute plasma cortisol levels. Plasma cortisol were objectively measured and considered biological, yet the heritability ranged from 0 to 6% depending on the sample and method.

Perhaps more stable measures of cortisol, which are not as responsive to acute situational contexts, are needed to improve the so far somewhat inconsistent study of cortisol genetics. Hair cortisol samples are an interesting candidate for long-term profiles. However, as shown in chapter V.1, cortisol measurements in hair are biased by the darkness of hair color. This makes genome-wide analyses difficult, as researchers have to exclude SNPs in linkage disequilibrium with hair color coding genes. In addition, hair samples may not represent months-long exposures as is often claimed, but may be biased by acute events.<sup>9</sup> Finally, it is not a given that basal levels of cortisol are relevant for child psychiatry. In unpublished analyses, we associated hair cortisol with general and specific psychopathology factors in a structural equation model, similar to the white matter integrity models. In these analyses we were not able to detect any associations between hair cortisol and child psychopathology. Perhaps cortisol reactivity to stress is more relevant to psychopathology than basal levels, with several studies finding associations between cortisol reactivity to stressors and concurrent or later psychopathology.<sup>10,11</sup>

## Methodological considerations

### Method factors in multi-informant/method models

The use of multiple informants, methods and assessment timepoints is important to account for shared method bias and to increase precision. Shared method bias occurs when predictor and outcome are measured with the same method, in the case of questionnaire data with the same questionnaire or the same informant. This poses the danger that a bias inherent to the same questionnaire/informant is present for both the predictor and outcome, creating spurious associations, which are actually just an indication of shared bias. An example would be the association of child and parent psychopathology using parental ratings for both.<sup>12</sup> This problem can also affect latent factor models, as they are fitted to the covariance of two variables, which may be inflated due to shared variance biases.

However, while the theoretical advantages of using multiple informants are quite clear, the actual implementation is much less so. A simple approach is to take the average of two or more informants, however, this ignores the possibility that some items or scales are affected more by bias than others, possible resulting in overadjustment for some items and underadjustment for others. Another possibility is the use of latent factors, as used in this dissertation with two variations. In chapter III.2 we introduce latent informant factors, which load on an all items rated by the same informant. In

chapter III.3 and III.5 however, we use context factors which load on items stemming from a single rating context, i.e. from the same informant, timepoint and same instrument. These method factors explain a substantial symptom variance with either approach, typically even more than the actual psychopathology factors. For example, in the white matter integrity model, maternal ratings at age 6 had loadings between 0.39 and 0.59 on the general psychopathology factor. The loadings on the rating context factor were between 0.32 and 0.69. Interestingly, in both cases the subscale summarizing “other” items had the highest loading, suggesting that ratings on this scale are at the same time representative of general psychopathology factor but also context bias. The comparable loading between psychopathology and method factors highlights the presence of a large disagreement between raters, timepoints and instruments. It should be noted that this disagreement does not necessarily reflect the unreliability of the rater, as different raters may have different unique insights or some ages and instruments may be better suited to assess psychopathology. Therefore great care must be taken in how rating biases are corrected, as different method factors can have large impacts on the model. The models in chapter III.2 and III.3 can be easily compared as they mostly differ in how the method factors are defined. Defining the method factors on the informant level caused a bigger suppression of loadings on the psychopathology factors than defining the method factors on more narrower rating contexts (ratings from one informant with one instrument at one time-point). The first approach has a more strict control for informant effects, as these are expected to persist between instruments and time points. However, because most items in the age 6-8 years models used in III.2 and III.3 are based on mother report, there was a risk that the mother factor forms a competing general psychopathology factor instead of just covering biases. It would thus strengthen this multi-informant approach, if the number of items from the different informants were more balanced. This can be difficult though, as not all informant can rate all items. Alternatively, it may be also possible to designate the mother report as reference method, as proposed in CT-C(M-1) methods<sup>13</sup>, and leave out the method factor for maternally rated items. All other method factors would then estimate the disagreement with the mother ratings. However, while there is an argument to be made that mothers spend most time with the children in many families and thus may have the best insight, mother ratings suffer from biases as well and therefore require correction.

Interestingly, the choice of method factor approach had great impact on the covariance between the specific factors. Using informant factors, the covariance is negligible, but with context factors the correlation is substantial. It is unclear to me where this difference stems from. One possibility is that by comprising all items from one rater across multiple instruments and time points, the method factor become a competing general psychopathology factor, reflecting the unique insight of an informant. Additionally, it may be that informant based method factors reflect the informant’s general psychopathology, which in case of parental rating is related to the child’s general psychopathology. Both situations imply that the specific psychopathology factors potentially are

more strictly defined in this case than if the method factors are based on single rating contexts.

In the future, it would be advantageous to have separate method factors for informants, instrument and timepoint. However, such a model with effectively four levels is likely too complex and more prone to overfitting and convergence problems. However, with a large enough sample size and with an exhaustive study design which covers all combinations of rater, instrument and time-point this may be possible. Such a design would be beneficial even with simpler method adjustments. It may be also interesting to relate the method factors to biological correlates to examine to which degree they may contain substantive psychopathology information, which is not simply bias. For example, one could explore whether parental genetic risk for psychopathology explains the parents' method factors and whether the genetic risk was transmitted to the children.

### **Effect sizes in epidemiology**

This dissertation presents many statistically significant results, but generally low effect sizes. For instance, the top CpG site associated with ADHD explained 0.25% of the variance in the Generation R Study, global white matter explains about 0.49% of the general psychopathology variance, and maternal age at most 0.66% of externalizing behavior. At first glance, one could therefore conclude that this thesis largely identified correlates of psychopathology with little relevance and these factors are not worth considering further in the quest to explain the development of child psychopathology. The notion, that the medical field should disregard results which are statistically significant but have low effect sizes makes sense in the context of randomized controlled trials testing intervention. If a medical doctor needs to decide which treatment to prescribe, it is important to choose one which provides a meaningful change to the patient. Otherwise, the costs and risk associated with any intervention may outweigh the benefits. However, in the case of epidemiological research, where the focus is on etiology, it is much less clear what constitutes the threshold for relevance. There are no risks to weigh against and no decision of one treatment against another: many different risk factors could and probably do act jointly to cause psychiatric symptoms. This phenomenon can be demonstrated and is well accepted in psychiatric genetics. The top SNP in the GWAS of total psychiatric symptoms alone only explains 0.09% of the variance. However, the overall SNP heritability is estimated at 8.4%. This is comparable to other GWASs of continuous traits, such as depression (top SNP explained 0.03% with a SNP heritability of 4.7%) or neuroticism (top SNP explained 0.04% with a SNP heritability of 9.1%).<sup>14</sup> Thus while, the effects of single SNPs appear small, the effects add up to meaningful proportions.

Could the same logic be applied to non-genetic data? On the one hand, genetic variables, such as variation in SNPs, do not suffer from reverse causality biases, and confounding biases are assumed to be more limited. However, there are three notable

exceptions: population stratification, gene-environment correlations and collider bias. Population stratification occurs, when environmental influences on a phenotype differ between participants of different ancestries. Those genetic variants which differ in frequency between ancestries then are a marker for environmental differences and are thus only spuriously associated with the phenotype. Several methods are applied to adjust for population stratification, including stratified analysis, principal component adjustment or linear mixed models. Population stratification can be seen as a specific example of gene-environment correlation, though many other examples can be thought of. For instance, as parents and child share genotype, a child's genotype may also be a marker for the parenting abilities of their parents, and thus a marker for an environmental influence.<sup>15</sup> Finally, collider bias can occur, if selections are occurring depending on both the phenotype levels and genotype. This could e.g. occur for those participants who have both the highest genetic risk for psychopathology and higher levels of psychopathology, as they are more likely to not participate as individuals only having genetic risk or actual high levels, potentially because they have the highest burden of adverse conditions.<sup>16</sup>

Non-genetic variables come with an even higher risk of confounding in epidemiological studies and often contain the effects of multiple factors. Also, reverse causality is often a challenge, which can be ruled out in genetic studies. These issues are controlled for in experimental studies, so these could provide some orientation with regard to what effect sizes to expect. Previous reviews of experimental psychology studies found that the average correlation is 0.21 (SD = 0.15) between experimental condition and psychological outcome, which is equivalent to an explained variance 4.41%. When interpreting these effect sizes, one should keep in mind that experimental studies test acute effects with experimental conditions chosen strong enough (and as a consequence arguably become unrealistic and non-generalizable) to elicit meaningful changes. This is in contrast to observational studies in general populations, where the determinants have naturally occurring distributions and outcomes are often the effects of long-term exposures. In addition, in order to control for confounding variables, determinants are adjusted for potential confounders. In the process, only the independent effects remain, but some of the effects shared with the confounder may represent true effects, which are lost. Thus, the more adjusted an analysis in an observational study is, the lower the expected effect size. This is a problem that experimental studies do not face. All in all, we therefore expect lower effect sizes in observational studies than in experimental studies. Given that experimental studies in psychology typically tend to have low effect sizes, it would be probably unlikely that we would find single causes of psychiatric symptoms explaining the majority of cases. It follows that it is necessary for non-genetic studies to identify as many determinants as possible, regardless of their effect size, and investigate their joint effect. For example, in the global white matter model of general psychopathology adjusted for maternal psychopathology, the most independently associated variable was maternal interpersonal sensitivity, explaining 3.6% of the gen-

eral psychopathology factor score variance. However, the whole model explains 28.4% (10-fold cross-validated  $R^2$ , with 100 repetitions), illustrating that the aggregation of relatively few variables (19) can explain a quarter of general psychopathology. Finally, intervention strategies based on epidemiological insights may be optimized to achieve larger effect sizes compared to the naturally occurring exposure. In conclusion, research should expect that properly confounding controlled variables will have small effect sizes in epidemiological studies. However, these should not be ignored, but jointly analyzed for further insights into the etiology of psychiatric disorders.

### **Measurement and relational invariance of the general psychopathology factor**

An important question for any study is to which populations research findings are generalizable. For instance, many of the featured studies use samples with either participants of only one genetic ancestry (e.g. European ancestry) or from only one geographical location (e.g. Rotterdam). Certainly, the most convenient scenario is when findings would be applicable to any population and are generally the same no matter of group belonging. For instance, if we could demonstrate that general psychopathology factors can be modeled the same way no matter the ancestry of participants or their socioeconomic status, then one could argue with more confidence that the findings would apply to other parts of the world, with different ethnic compositions or financial wealth. However, generalizability can be wanting in at least two ways in psychiatric epidemiology. First, different groups may express the same levels of psychopathology differently (lack of measurement invariance). It is conceivable, that the concept of general psychopathology applies to both girls and boys, but that its expression may be different. For example, boys compared to girls with the same amount of general psychopathology may show more aggressiveness (higher loadings of aggressiveness on general psychopathology) or show more aggression independent of the levels of any psychopathology factor (higher intercept of aggression). If this is the case it would be invalid to associate the general psychopathology factor across genders with any determinant or outcome, as the individual's psychopathology factors would represent different symptom sets depending on gender. In this dissertation we tested the measurement invariance of general psychopathology factors across sex, ancestry and socioeconomic status. We found that across all these groups (as defined, for example, by sex) the psychopathology factors were invariant with respect to intercepts and loadings (strong invariance), but not with regard to the residuals. Since strong invariance is a sufficient condition to associate the factors with other variables across groups, we did not discuss this issue further initially, however, I find it worthwhile diving into this matter here.

Strong invariance implies, that the basic structure of psychopathology holds across groups, and that the predicted symptom scores are the same for children with the same levels of the psychopathology factors. However, for sex and ancestry, the residual variance was not equal, which means that the factors explain different amounts of variance

depending on the sex or ancestry. Specifically, particularly externalizing and attention symptom scores are better explained in girls than boys and almost all psychiatric symptom scores have a higher variance explained in children with non-European ancestry. This in turns implies that the correlates presented in this thesis differ in the proportion of explained variance depending on the specific symptoms, the sex of the child and their ancestry. With regard to sex, a possible interpretation is that in boys externalizing symptoms are more specific, e.g. boys are more likely to show aggression without other accompanying problems, even other externalizing symptoms. This interpretation may also hold in case of observer biases, i.e. if boys do not have higher symptoms but were nevertheless rated higher, because symptoms were easier to recognize than in girls. Another possibility is, that there is another latent psychopathology factor present only in boys, though it is hard to imagine what this factor would be. In the case of ancestry, the explained variance for most subscales was higher in children with non-European ancestry. How could it be that children with European ancestry have higher proportions of variance unexplained by the tested psychopathology factors? Again, it is difficult to speculate how the missing explained proportion came about. Since the majority of the psychiatric assessments are based on parental and self report, cultural differences might play a role. Specifically, individual symptoms in children of European (mostly Dutch) ancestry could be rated higher independent of other symptoms. Rather than cultural differences per se, minority status may also play a role. Perhaps children facing more challenges due to being a minority are more likely to suffer from broader psychopathology symptoms rather than single symptoms, and thus the general psychopathology factor models fit better. Socioeconomic status as defined by maternal education does not appear to explain this discrepancy. Children with mothers from higher and lower educational background did not differ in any model parameters, including the residual error. Thus, the general psychopathology factor model is strictly invariant with respect to maternal education. Finally, the higher explained variance may be the result of higher variance in ratings of children with non-European background. It may be easier for the model to explain the co-occurrence of symptoms if there is a greater diversity in scores of children with and without psychiatric problems.

In summary, while it appears that the general and specific psychopathology factors show strong invariance across many of the tested groups, the models introduced in this thesis differ in their explanatory power. Further research is therefore needed to explore why the residual variance differs and how this could be remedied.

## **Clinical Implications**

When drafting the chapters and presenting the results, a frequent question raised by reviewers and co-authors concerned the possible clinical implications of the findings. Indeed, at first glance the latent factors general psychopathology and the specific psychopathology factors are not observable, but abstract concepts. How would a clinician assess a patient's general or specific psychopathology levels? In the case of general



psychopathology, a good approximation would be the sum of any symptoms. However, in the case of specific psychopathology, it is more challenging to define a clinical picture. A child displaying specific psychopathology has a set of symptoms from one domain independent of their general level of psychopathology. However, independence here does not mean that both cannot co-occur. It is likely that a specific symptom occurs as the result of general and specific effects. Thus from a clinical assessment perspective, the proposed bifactor models are not directly applicable. However, these models do give valuable insights into the etiology of psychopathology and thus clues to the best prevention and treatment strategies, and may aid prediction of later psychiatric disorders. For instance, in our and other studies neuroticism or negative affect reactivity was consistently and strongly related to general psychopathology. Neuroticism is a personality trait and may be difficult to change, however, if one were to successfully change this trait, the predicted effects would be immensely helpful, as they should affect a broad spectrum of psychiatric symptoms. Alternatively, environmental stressors affecting mood would have to be eliminated as much as possible to reduce the effects of high neuroticism. On a biological level, the genome-wide and epigenome-wide association studies suggest targeting DMWD, ERC2 or CREB5 expression or its gene products, however, experimental research is needed to first confirm the causal role of these genes. Next to providing leads for intervention, biological studies of general psychopathology may be useful prediction. The combination of genetic scores, based on both general and specific psychiatric GWASs, polygenic scores taking into account gene-environment interactions, as well as polyepigenetic scores based on DNA methylation should provide meaningful predictions of psychiatric risk. The predictive power certainly will improve with increasing sample sizes of epigenetic studies.

### **Future research and the need for higher sample sizes**

In the previous chapters specific recommendations for future research have been already discussed. Thus here I only present general observations. Increasing sample size in research of child psychiatric disorders is an important, yet challenging aspect. The need for an increase in sample size is obvious in situations where there is not enough power to detect an effect in the first place. For example, we identified three genes in the genome-wide association study of a total child psychiatric sum score, but likely many more genes are important in regulating symptoms. Likely variation at more specific DNA methylation sites were associated with ADHD, also at school-age. While at school age no methylation levels were genome-wide significant for any CpG site, the regression coefficients did show some correlation with regression coefficients at birth in independent samples. It is therefore likely that DNA methylation at school-age is also associated with ADHD, but we do not yet have the power to detect these effects. One may argue, that any SNP or probe with meaningful effects would have already been detected with current sample sizes and that an increase would only reveal unimportant loci. However, as discussed earlier, even very small effects seem to add up to substantial

magnitudes, thus detecting more variants and having more precise estimates of their association is of great importance.

An increase in sample size may also be important to evaluate associations for which there is enough power, but where the magnitude of the association is uncertain, such as for SNP heritability estimates presented in this study. For instance, while we can rule out large contributions of genetics towards cortisol, some genetic contribution is expected, but it is difficult to determine the precise magnitude. This is especially true for the repeated cortisol measures, for which we only had a low sample size available.

Finally, a higher sample size would allow for more sophisticated, detailed and better adjusted analysis. Even for those studies, where we had enough power to answer the main research question and have relatively precise estimates, several follow up analyses were not possible. A good example is the study on white matter. While we detected an association between white matter integrity and general psychopathology, there also appeared to be an interaction with sex. The effect appeared to be stronger in boys, but the difference was not statistically significant. It would be interesting to follow up in a larger study or meta-analysis, whether the observed sex difference was just chance or a robust effect. Also a longitudinal design examining changes in general psychopathology and white matter integrity throughout childhood would be highly interesting to determine the directionality of effect. However, such analyses may be difficult given the complexity of the bifactor model and the need to partition symptom variance into general, specific and method proportions. Convergence can become easily a problem at lower sample sizes and if there are not enough items. These problems, however, can be avoided by combining several assessment waves as done here, but for longitudinal modeling, general psychopathology measures per assessment waves are necessary, which require very large sample sizes. Furthermore, it would be also interesting to jointly analyze genetics and DNA methylation. This could help separate environmental from genetic mediation effects. In addition, interaction between the genome and methylome could be accounted for. But again, such conditional and interaction analyses need larger sample sizes than what is typically available now.

Finally, as hopefully demonstrated in this thesis, it is advantageous to analyze several phenotypes simultaneously to be able to differentiate general from specific effects and to increase precision. However, the more measures one attempts to analyze, the higher the chance that at least some data is missing. While missing data techniques can to some extent remedy this problem, higher sample sizes than in single phenotype analyses are still beneficial to account for the missingness.

## Summary

In this thesis we investigated general psychopathology factors in school-aged children in relation to various biological correlates. We consistently observed across different cohorts, that a substantial proportion of variation in psychiatric symptoms can be attributed to general psychopathology effects. This observation is unlikely to be at-

tributable to informant or other rating context biases. The general psychopathology factor appears to be partly heritable, with single nucleotide polymorphisms playing a central role. One locus was identified to be associated with general psychopathology, the myotonic dystrophy cluster. Overall those genes, that are expressed in the brain, particularly in the limbic regions, appear especially important in the genetics of general psychopathology. On a neural level, more white matter across the brain is associated with lower levels of general psychopathology. At the same time, more white matter across the brain is associated with more levels of specific externalizing levels. Additionally, we observed that DNA methylation at birth is associated with ADHD symptoms in school-age and that higher maternal and paternal age is associated with less externalizing problems. Finally, we observed low SNP heritability of acute cortisol levels, but also highlight that hair cortisol levels may be biased by hair pigmentation.

## REFERENCES

1. Lahey BB, Van Hulle CA, Singh AL, Waldman ID, Rathouz PJ. Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Arch Gen Psychiatry*. 2011;68(2):181-189. doi:10.1001/archgenpsychiatry.2010.192
2. Caspi A, Houts RM, Belsky DW, et al. The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin Psychol Sci*. 2014;2(2):119-137. doi:10.1177/2167702613497473
3. Laceulle OM, Vollebergh W a. M, Ormel J. The Structure of Psychopathology in Adolescence: Replication of a General Psychopathology Factor in the TRAILS Study. *Clin Psychol Sci*. 2015:1-11. doi:10.1177/2167702614560750
4. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is There a General Factor of Prevalent Psychopathology During Adulthood? *J Abnorm Psychol*. 2012;121(4):971-977. doi:10.1037/a0028355
5. Bonifay W, Cai L. On the Complexity of Item Response Theory Models. *Multivariate Behav Res*. 2017;0(0):1-20. doi:10.1080/00273171.2017.1309262
6. Nuijten MMB, Deserno MMK, Cramer AOJ, Borsboom D, ... AC-C, 2016 undefined. Mental disorders as complex networks: An introduction and overview of a network approach to psychopathology. *Clin Neuropsychiatry*. 2016;13(4/5):68-76. <https://www.narcis.nl/publication/RecordID/oi:dare.uva.nl:publications%2F5eb5a18e-1af5-429a-8ef0-f315df2bccad>. Accessed January 7, 2019.
7. Walton E, Pingault J, Gaunt TR, et al. Epigenetic profiling of ADHD symptoms trajectories: A prospective, methylome-wide study Esther. 44(0).
8. Flint J, Munafò MR. The endophenotype concept in psychiatric genetics. *Psychol Med*. 2007;37(2):163-180. doi:10.1017/S0033291706008750
9. Kalliokoski O, Jellestad FK, Murison R. Are we overestimating the utility of hair glucocorticoids? A systematic review exploring the empirical evidence supporting hair glucocorticoids as a measure of stress. *bioRxiv*. July 2018:375667. doi:10.1101/375667
10. Saxbe DE, Margolin G, Spies Shapiro LA, Baucom BR. Does Dampened Physiological Reactivity Protect Youth in Aggressive Family Environments? *Child Dev*. 2012;83(3):821-830. doi:10.1111/j.1467-8624.2012.01752.x
11. Laurent HK, Gilliam KS, Wright DB, Fisher PA. Stress Responses : Evidence of Developmental Stress Sensitization. *J Abnorm Psychol*. 2015;124(1):68-79.
12. Ringoot AP, Tiemeier H, Jaddoe VWV, et al. Parental depression and child well-being: young children's self-reports helped addressing biases in parent reports. *J Clin Epidemiol*. 2015. doi:10.1016/j.jclinepi.2015.03.009
13. Papa LA, Litson K, Lockhart G, Chassin L, Geiser C. Analyzing statistical mediation with multiple informants: A new approach with an application in clinical psychology. *Front Psychol*. 2015;6(NOV):1-12. doi:10.3389/fpsyg.2015.01674
14. Okbay A, Baselmans BML, De Neve JE, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*. 2016;48(6):624-633. doi:10.1038/ng.3552
15. Knafo A, Jaffee SR. Gene-environment correlation in developmental psychopathology. *Dev*

Psychopathol. 2013;25(1):1-6. doi:10.1017/S0954579412000855

16. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol.* 2017;47(1):226-235.





Chapter VII

**Summary/Samenvatting**

## SUMMARY

Co-occurrence of mental disorders is widespread and studies have identified a general psychopathology factor reflecting vulnerability to experience a range of psychiatric problems (**Chapter II**). However, the biological mechanisms underlying the co-occurrence of child psychiatric symptoms remain unclear. The main question of this dissertation was: which biological factors are associated with child psychopathology in general and which biological factors are specific to certain psychopathology domains?

In **Chapter III.A** we examined the contributions of maternal and paternal age on offspring externalizing and internalizing problems, this study analyzed problem behaviors at age 10-12 years from four Dutch population-based cohorts (N = 32,892). There was evidence of a robust negative linear relation between parental age and externalizing problems as reported by parents. Parental age had limited to no association with internalizing problems. Thus, in this large population-based study, either a beneficial or no effect of advanced parenthood on child problem behavior was observed.

In **Chapter III.B** we developed a model of childhood psychopathology that separates the unique and shared contribution of individual psychological symptoms into specific internalising, externalising and general psychopathology factors and assess how these general and specific factors predict long-term outcomes concerning criminal behaviour, academic achievement and affective symptoms. Child psychopathology was assessed repeatedly using a range of diagnostic and questionnaire-based measures, and multiple informants. As hypothesized, general psychopathology factor scores were predictive of all outcomes of later functioning, while internalising factor scores specifically predicted later internalising outcomes. Externalising factor scores, capturing variance not shared by any other psychological symptoms, were not predictive of later outcomes.

In **Chapter III.C** We defined a multi-informant general psychopathology factor in school-aged children and estimated its SNP heritability. The goal was to test the hypothesis that child behavioral and emotional problems are under the influence of highly pleiotropic common autosomal genetic variants that non-specifically increase the risk for different dimensions of psychopathology. Children were repeatedly assessed between ages 6-8 years. Child behavior problems were reported by parents, teachers and children. The general psychopathology factor showed a significant SNP heritability of 38%.

To identify the specific genetic variants underlying global childhood psychopathology, we then performed a genome-wide association study of a total psychiatric problem score (**Chapter III.D**). We analyzed 8,804,648 commonSNPs in 29,446 school-aged children from 16 population-based cohorts participating in the Early Genetics and



Lifecourse Epidemiology (EAGLE) consortium. Gene-based analyses revealed, that the myotonic dystrophy (DM1) gene cluster, previously implicated in neurodevelopment, was associated with the total psychiatric problem score. No individual SNP reached genome-wide significance.

In **chapter III.E** We tested the hypothesis that lower overall white matter microstructure is associated with higher levels of the general psychopathology factor in children and less with specific factors. Global white matter microstructure at age 10 years was related to general and specific psychopathology factors. These factors were estimated using a latent bifactor model with multiple informants and instruments between ages 6-10 years in 3030 children. Higher levels of global white matter were associated with lower general psychopathology. In contrast, more white matter microstructure predicted an increase of specific externalizing factor levels. No association was found with the specific internalizing and specific attention factor.

In **Chapter IV** We performed an epigenome-wide association study within the Pregnancy And Childhood Epigenetics (PACE) Consortium to identify DNA methylation sites associated with ADHD symptoms. As DNA methylation changes over time, we used two assessment periods: birth and school-age. We examined associations of DNA methylation in cord blood with repeatedly assessed ADHD symptoms (age range 4-15 years) in 2477 children from five cohorts and DNA methylation at school-age (age 7-9 years) with concurrent ADHD symptoms (age 7-11 years) in 2374 children from ten cohorts. We identified 9 probes at birth that were associated with later ADHD symptoms. In contrast, no probes reached genome-wide significance when ADHD was associated with school-age DNA methylation.

Cortisol concentrations in hair are used to create stress hormone profiles spanning months. However, this method may be biased by hair pigmentation. We tested the hypothesis that hair pigmentation gene variants are associated with varying cortisol levels independent of genetic ancestry (Chapter V.A). Hormone concentrations and genotype were measured in 1674 children from the Generation R cohort at age 6. We computed a polygenic score of hair color based on 9 single nucleotide polymorphisms. This score was used to predict hair cortisol concentrations, adjusted for genetic ancestry, sex, age and corticosteroid use. A higher polygenic score (darker hair) was associated with higher cortisol levels. This suggests that variation in hair cortisol concentrations is partly explained by local hair effects.

In **Chapter V.B** Our aim was to estimate the amount of cortisol variance explained by common single nucleotide polymorphisms, i.e. the SNP heritability. We analyzed morning plasma ( $n=5,705$ ) and saliva levels ( $n=1,717$ ), as well as diurnal saliva levels ( $n=1,541$ ), in the Rotterdam Study and data from the CORNET consortium on morning plasma cor-

tisol (n=12,597) and saliva cortisol (n=7,703). No significant SNP heritability was detected for any cortisol measure, sample or analysis approach. Point estimates ranged from 0% to 9%. Morning plasma cortisol in the CORNET cohorts, the sample with the most power, had a 6% SNP heritability. The results consistently suggest a low SNP heritability of these acute and short-term measures of cortisol.

Finally, **Chapter VI** discusses overarching findings, methodological consideration and the implications of the study results.

## SAMENVATTING

Het gelijktijdig voorkomen van psychische stoornissen is wijdverbreid en studies hebben een algemene psychopathologische factor geïdentificeerd die de kwetsbaarheid voor een reeks aan psychiatrische problemen weerspiegelt (**Hoofdstuk II**). De biologische mechanismen die ten grondslag liggen aan de co-aanwezigheid van kinderpsychiatrische symptomen blijven echter onduidelijk. De hoofdvraag van dit proefschrift was: welke biologische factoren zijn geassocieerd met kinderpsychopathologie in het algemeen en welke biologische factoren zijn specifiek voor bepaalde psychopathologische domeinen?

In **Hoofdstuk III.A** onderzochten we de bijdragen van de leeftijd van moeders en vaders aan de externaliserende en internaliserende problemen van nakomelingen. Deze studie analyseerde problematisch gedrag op de leeftijd van 10-12 jaar van vier Nederlandse algemene bevolkings cohorten (N = 32.892). Er was bewijs voor een robuust negatieve lineaire associatie tussen leeftijd van de ouders en externaliserende problemen zoals gerapporteerd door ouders. De leeftijd van de ouders was niet geassocieerd met internaliserende problemen. Dus, in dit grote algemene bevolkingsonderzoek, werd ofwel een voordelig of geen effect van vergevorderd ouderschap op het gedrag van het kind-probleem waargenomen.

In **Hoofdstuk III.B** hebben we een model ontwikkeld voor kinderpsychopathologie dat de unieke en gedeelde bijdrage van individuele psychologische symptomen scheidt in specifieke internaliserende, externaliserende en algemene psychopathologische factoren en evalueert hoe deze algemene en specifieke factoren psychiatrische diagnoses, crimineel gedrag en academisch prestatie in late adolescentie en jonge volwassenheid voorspellen. Kinderpsychopathologie werd herhaaldelijk beoordeeld aan de hand van een reeks diagnostische en op vragenlijst gebaseerde metingen verkregen van meerdere informanten. Zoals verondersteld, waren de algemene psychopathologie factorscores voorspellend voor alle uitkomsten van het latere functioneren, terwijl internaliserende factorscores specifiek latere internaliserende uitkomsten voorspelden. Externaliserende factorscores, waarin variantie werd gevangen die niet werd gedeeld door andere psychologische symptomen, waren niet voorspellend voor latere uitkomsten.

In **Hoofdstuk III.C** hebben we een multi-informante algemene psychopathologische factor gedefinieerd in schoolgaande kinderen en de SNP-erfelijkheid ervan geschat. Het doel was om de hypothese te testen dat gedrags- en emotionele problemen bij kinderen onder invloed zijn van zeer pleiotrope vaak voorkomende autosomale genetische varianten die niet-specifiek het risico verhogen voor verschillende dimensies van psychopathologie. Kinderen werden herhaaldelijk beoordeeld tijdens de leeftijd 6-8 jaar.

Problemen met het gedrag van het kind werden gemeld door ouders, leerkrachten en kinderen. De algemene psychopathologische factor vertoonde een significante SNP-erfelijkheidsgraad van 38%.

Om de specifieke genetische varianten te identificeren die ten grondslag liggen aan algemene psychopathologie bij kinderen, hebben we vervolgens een genomwijde associatietudie uitgevoerd over een totaalscore van psychiatrisch problemen (**Hoofdstuk III.D**). We analyseerden 8.804.648 vaak voorkomende SNP's in 29.446 schoolgaande kinderen uit 16 populatie-gebaseerde cohorten die deelnamen in het Early Genetics and Lifecourse Epidemiology (EAGLE) consortium. Analyse op basis van genen onthulde dat het myotone dystrofie (DM1) -cluster, welke betrokken is in de neurologische ontwikkeling, geassocieerd was met de totaalscore van psychiatrische problemen. Geen enkele individuele SNP bereikte genomwijde significantie.

In **Hoofdstuk III.E** hebben we de hypothese getest dat een lagere globale witte stof microstructuur in kinderen geassocieerd is met hogere niveaus van de algemene psychopathologische factor en minder met specifieke factoren. Globale witte stof microstructuur op de leeftijd van 10 jaar werd gerelateerd aan algemene en specifieke psychopathologische factoren. Deze factoren werden geschat met behulp van een latent bifactor model met meerdere informant en instrumenten op de leeftijd van 6-10 jaar bij 3030 kinderen. Hogere niveaus van globale witte stof waren geassocieerd met lagere algemene psychopathologie. Daarentegen voorspelde een hogere microstructuur van witte stof een toename van specifieke factoren voor externaliserende factoren. Er werd geen associatie gevonden met de specifieke internaliserende of specifieke aandachtsfactor.

In **Hoofdstuk IV** voerden we een epigenome-wide association study uit in het Pregnancy And Childhood Epigenetics (PACE) Consortium om DNA-methylatie locaties geassocieerd met ADHD-symptomen te identificeren. Omdat DNA-methylatie verandert over de tijd, hebben we twee beoordelingsperiodes gebruikt: geboorte en schoolleeftijd. We onderzochten associaties van DNA-methylatie in navelstrengbloed met herhaaldelijk vastgestelde ADHD-symptomen (leeftijd 4-15 jaar) bij 2477 kinderen uit vijf cohorten en DNA-methylatie op schoolleeftijd (leeftijd 7-9 jaar) met gelijktijdige ADHD-symptomen (leeftijd 7- 11 jaar) in 2374 kinderen uit tien cohorten. We identificeerden 9 locaties bij de geboorte die geassocieerd waren met latere ADHD-symptomen. Daarentegen bereikten geen probes een genomwijde significantie wanneer ADHD werd geassocieerd met DNA-methylatie op schoolleeftijd.

Cortisolconcentraties in het haar worden gebruikt om stresshormoonprofielen te creëren die voorgaande maanden representeren. Deze methode kan echter beïnvloed worden door haarpigmentatie. We testten de hypothese dat varianten van haarpigmen-

tatiegenen geassocieerd zijn met cortisolspiegels onafhankelijk van genetische afstamming (**Hoofdstuk V.A**). Hormoonconcentraties en genotype werden gemeten in 1674 kinderen uit het Generation R-cohort op de leeftijd van 6. We berekenden een polygene score van haarkleur op basis van 9 single-nucleotide polymorfismen. Deze score werd gebruikt om haarcortisolconcentraties te voorspellen, gecorrigeerd voor genetische afkomst, geslacht, leeftijd en gebruik van corticosteroïden. Een hogere polygene score (donkerder haar) was geassocieerd met hogere cortisolspiegels. Dit suggereert dat de variantie in cortisolconcentraties in het haar gedeeltelijk kan worden verklaard door lokale haareffecten.

In **Hoofdstuk V.B**. was ons doel om de hoeveelheid cortisolvariantie te schatten, verklaard door algemene single-nucleotide polymorfismen, dat wil zeggen de SNP erfelijkheid. We analyseerden ochtendplasma ( $n = 5.705$ ) en speekspiegels ( $n = 1.717$ ), evenals dagelijkse speekselwaarden ( $n = 1.541$ ), in de Rotterdam Study en ochtendplasma cortisol ( $n = 12.597$ ) en speeksel cortisol ( $n = 7.703$ ) in het CORNET-consortium. Er werd geen significante SNP erfelijkheid gedetecteerd onafhankelijk van cortisol-meting, populatie of analysebenadering. Puntchattingen varieerden van 0% tot 9%. Ochtendplasma cortisol in de CORNET cohorten, de populatie met de meeste power, had een SNP erfelijkheid van 6%. Deze consistente resultaten suggereren een lage SNP-erfelijkheid van deze acute en kortdurende metingen van cortisol.

Ten slotte bespreekt **Hoofdstuk VI** overkoepelende bevindingen, methodologische overwegingen en de implicaties van de studieresultaten.





Chapter VIII

APPENDIX

**Acknowledgments**

**Publications and Manuscripts**

**Portfolio**

**Words of thanks**





## ACKNOWLEDGMENTS

GenR is conducted by the Erasmus Medical Center in close collaboration with the Erasmus University Rotterdam, Faculty of Social Sciences, the Municipal Health Service Rotterdam area, and the Stichting Trombosedienst and Artsenlaboratorium Rijnmond (STAR), Rotterdam. We gratefully acknowledge the contribution of general practitioners, hospitals, midwives and pharmacies in Rotterdam. The Generation R Study is made possible by financial support from: Erasmus Medical Center, Rotterdam, and the Netherlands Organization for Health Research and Development (ZonMw).

The presented work and studies were performed at the Department of Child and Adolescent Psychiatry/Psychology of the Erasmus University Medical Center – Sophia Children’s Hospital and supported by a grant of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant No. 024.001.003, Consortium on Individual Development). The author was additionally supported by a Canadian Institutes of Health Research team grant.

## PUBLICATIONS AND MANUSCRIPTS

### Published

Neumann, A., Pappa, I., Lahey, B. B., Verhulst, F. C., Medina-Gomez, C., Jaddoe, V. W., ... & Tiemeier, H. (2016). Single nucleotide polymorphism heritability of a general psychopathology factor in children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(12), 1038-1045. (Chapter III.C)

Neumann, A., Noppe, G., Liu, F., Kayser, M., Verhulst, F. C., Jaddoe, V. W., ... & Tiemeier, H. (2017). Predicting hair cortisol levels with hair pigmentation genes: a possible hair pigmentation bias. *Scientific reports*, 7(1), 8529. (Chapter V.A)

Neumann, A., Direk, N., Crawford, A. A., Mirza, S., Adams, H., Bolton, J., ... & Milaneschi, Y. (2017). The low single nucleotide polymorphism heritability of plasma and saliva cortisol levels. *Psychoneuroendocrinology*, 85, 88-95. (Chapter V.B)

Viuff, A. C., Sharp, G. C., Rai, D., Henriksen, T. B., Pedersen, L. H., Kyng, K. J., ... & Tiemeier, H. (2018). Maternal depression during pregnancy and cord blood DNA methylation: findings from the Avon Longitudinal Study of Parents and Children. *Translational psychiatry*, 8(1), 244.

Vojinovic, D., Adams, H. H., Jian, X., Yang, Q., Smith, A. V., Bis, J. C., ... & Saba, Y. (2018). Genome-wide association study of 23,500 individuals identifies 7 loci associated with brain ventricular volume. *Nature communications*, 9(1), 3945.

Cortes Hidalgo, A. P., Neumann, A., Bakermans-Kranenburg, M. J., Jaddoe, V. W., Rijlaarsdam, J., Verhulst, F. C., White, T., van IJzendoorn, M. H. and Tiemeier, H. (2019), Prenatal Maternal Stress and Child IQ. *Child Development*.

Cardenas, A., Faleschini, S., Hidalgo, A. C., Rifas-Shiman, S. L., Baccarelli, A. A., DeMeo, D. L., ... & El Marroun, H. (2019). Prenatal maternal antidepressants, anxiety, and depression and offspring DNA methylation: epigenome-wide associations at birth and persistence into early childhood. *Clinical Epigenetics*, 11(1), 56.

Bolhuis, K., Tiemeier, H., Jansen, P. R., Muetzel, R. L., Neumann, A., Hillegers, M. H., ... & White, T. (2019). Interaction of schizophrenia polygenic risk and cortisol level on pre-adolescent brain structure. *Psychoneuroendocrinology*, 101, 295-303.

## Accepted

Sallis, H., Szekely, E., Neumann, A., Jolicoeur-Martineau, A., van IJzendoorn, M., ... & Evans J. (2019). General psychopathology, internalising and externalising in children and functional outcomes in late adolescence. *Journal of Child Psychology and Psychiatry* (in press). (Chapter III.B)

Dunn, E.C., Nishimi, K., Neumann, A., Renaud, A., Cecil, C.A.M., ... & Tiemeier, H. (2019). Time-dependent effects of exposure to physical and sexual violence on psychopathology symptoms in late childhood: In search of sensitive periods in development. *Journal of the American Academy of Child & Adolescent Psychiatry* (in press).

Zondervan-Zwijnenburg, M.A.J., Veldkamp, S.A.M., Nelemans, S.A., Neumann, A., Barzeva, S., Branje, S. J. T., ... & Boomsma, D.I. (2019). Parental Age and Offspring Childhood Mental Health: A Multi-Cohort, Population-Based Investigation. *Child Development* (in press). (Chapter III.A)

## Submitted

Abdellaoui, A., Sandra, S., Sealock, J., Treur, J.L., Dennis, J.. Phenome-wide investigation of health outcomes associated with genetic predisposition to loneliness. *Human Molecular Genetic* (Submitted).

Diemer, E.W., Labrecque,, J.A., Neumann, A., Tiemeier, H., Swanson, S.A., Mendelian randomization approaches to the study of prenatal exposures: a systematic review. *International Journal of Epidemiology* (Submitted).

Neumann, A., Muetzel, R.L., Lahey, B.B., Bakermans-Kranenburg, M.J., van IJzendoorn, M.H., ... & Tiemeier, H. (2019). White matter microstructure and the general psychopathology factor in children (Submitted). (Chapter III.E)

## PhD Portfolio

Date: Wednesday, 5 March 2019

Name PhD student:

Alexander Neumann

Erasmus MC Department:

Child & Adolescent Psychiatry

Research School:

NIHES

PhD period:

Aug 2014 – Feb 2019

Promotors:

Henning Tiemeier

Marinus van IJzendoorn

Marian Bakermans-Kranenburg

	Year	ECTS
<b>1. PhD training</b>		
<b>MSc-program Genetic Epidemiology, NIHES:</b>		
Principles of Research in Medicine	2014	0.7
Principles of Genetic Epidemiology	2014	0.7
Genomics in Molecular Medicine	2014	1.4
Advances in Genomics Research	2014	0.4
Genome-Wide Association Analysis	2014	1.4
Study Design	2014	4.3
Genetic-Epidemiologic Research Methods	2014	5.1
Linux for Scientists	2014	0.6
SNP's and Human Diseases	2014	1.4
Biostatistical Methods I: Basic Principles	2015	5.7
Biostatistical Methods II: Classical Regression Models	2015	4.3
<b>Elective courses, NIHES:</b>		
Advances in GWAS	2015	1.4
Family-based Genetic Analysis	2015	1.4
An Introduction to the Analysis of Next-Generation Sequencing Data	2015	1.4
Topics in Meta-analysis	2015	0.7
Pharmaco-epidemiology	2015	0.7
Introduction to Bayesian Methods in Clinical Research	2015	1.4
Causal Mediation Analysis	2015	0.7
Social Epidemiology	2015	0.7
<b>Symposia, Conferences &amp; Workshops:</b>		
5 CID Meetings/Symposia , <i>Universiteit Utrecht</i>	2014-2018	1
EAGLE meeting, <i>Vrije Universiteit</i>	2015	0.2
Quantitative Genomics, <i>Wellcome Trust, London</i>	2015	0.2

World Congress of Psychiatric Genetics, Toronto	2015	1.4
Sophia Research Day (Oral Presentation)	2016	0.2
CID Lab Tour, <i>Universiteit Utrecht</i>	2016	0.2
CID Symposium, <i>Universiteit Utrecht</i> (Poster Presentation)	2016	0.2
KNICR Symposium, <i>Erasmus MC</i> (Oral Presentation)	2016	0.2
SRCD, Austin (Oral presentation)	2017	1.4
ISRCAP, <i>Universiteit van Amsterdam</i> (Oral presentation)	2017	1.4
Stress-NL, Amsterdam	2017	0.2
CID Meeting, <i>Universiteit Utrecht</i> (Oral presentation)	2018	0.2
Adult and Pediatric Life Support	2018	0.4
MRI safety and usage training	2016-2017	1
<b>2. General scientific activities</b>		
Supervising systematic review article writing for course: "Ontwikkeling van psychiatrische ziektebeelden"	2014-2015	0.5
Master Thesis Supervision: 2 students	2014-2015	6
Minor Thesis Supervision: 2 students	2015	0.5
Supervising systematic review article writing for course: "Ontwikkeling van psychiatrische ziektebeelden"	2015-2016	0.5
Master Thesis Supervision: 2 students	2016	6
Supervising systematic review article writing for course: "Ontwikkeling van psychiatrische ziektebeelden"	2016-2017	0.5
Peer review (JAACAP, JCPP)	2017	0.4
Minor Thesis Supervision: 2 students	2018	0.5
Medicine Student Internship supervision	2018	3
Peer review (JAACAP)	2018	0.2

## WORDS OF THANKS

This thesis described observations based on thousands of participants from all around the world. The studies performed here were made possible by countless researchers, staff and participants, who I wish I could thank all individually. Unfortunately, I will be able to only concentrate on those people, who directly helped me on a personal level to write this PhD thesis.

Naturally, the person most directly involved in a PhD student's life is the PhD promotor. I was very lucky to have Henning Tiemeier as promotor. He not only had an incredibly diverse knowledge on all things related to psychiatric epidemiology and related fields, but also always had good advice on how to conduct scientific research in general, as well as on how to survive politics and bureaucracy in science. I was always impressed how Henning was able to have detailed knowledge about all of his student's projects, given how many students Henning supervised at any given time. I have lots of good things to say about Henning, but the one aspect I appreciated the most is that I felt like I had freedom how I wanted to approach scientific questions, but at the same time he was also very decisive when I felt unsure or when he noticed that I was on a wrong path. In other words, I never felt neither lost nor restricted with Henning. I also have to thank my second and third promotor: Marinus van IJzendoorn and Marian Bakermans-Kranenburg. I appreciated their advice on all my projects and the chance to have an alternative perspective. I was also impressed, how I could send them a manuscript in the middle of the night and receive detailed comments back five minutes later.

While I was lucky to have very involved promotors with whom I had lots of contact, I also worked with many other colleagues who shaped my years as PhD student. In the first years, it was no doubt Irene Pappa, who had the most influence. Most things I know about genetic analyses are thanks to her, as she helped me immensely to enter the world of psychiatric genetics with lots of practical knowledge and guidance. I am also immensely happy, that Irene was one of my first friends in Rotterdam. I am very grateful for all the chats we had and her emotional support. Speaking of good friends, I was also very happy that Charlotte Cecil joined Generation R. She is the perfect combination of friend, colleague and supervisor. I have been learning a lot from her and am excited to see how the projects we started during the PhD period will turn out.

In general, the behavior and cognition group of Generation R is full of amazing people, who are pleasant to work with and inspire me. I would like to thank everybody in the behavior group, as everybody had some impact on me and helped me in writing of the thesis. I would like to especially thank Rosa Mulder for her help and friendship. While it is unfortunately not yet reflected in the authorship list, she had a big influence on most of the presented studies in this thesis. I am happy to see that finally the first studies in which we formally collaborated are being completed. Andrea Cortes, Ryan Muetzel and Jeremy Labreque, I enjoyed a lot talking with you about such fun topics like

measurement invariance, tract-based spatial statistics and sampling bias, though also happy that we could also talk about non-work related things such as dogs and marathons. Desi, you are a remarkable human being, as well.

Several people outside of the behavior and cognition group also had an impact on this thesis. For example, the people from the molecular epidemiology group, such as Carolina Medina-Gomez and Janine Felix, helped me learn a lot about genetics in general and gave lots of practical advice. Outside of Erasmus MC, I must thank Esther Walton, who helped me a lot with the EWAS of ADHD symptoms, as well as new projects. Her openness to work on challenging studies and her critical evaluation of my proposals has been crucial for these projects. I also want to thank Ashley Wazana and Eszter Szekely for the opportunity to join the DREAM BIG Consortium and for research visits at the Jewish General Hospital in Montreal. I want to thank them for always understanding, when I had to concentrate on my PhD thesis, for teaching me many new concepts, analysis approaches and perspectives. I am also grateful for the discussions and collaborations with Alexia Jolicoeur-Martineau, Jonathan Evans, Hannah Sallis, Rebecca Pearson, Erin Dunn and Yiwen Zhu. I also want to thank all co-authors and members of the EAGLE, PACE and CORNET consortia for their contributions and advice. Furthermore, I appreciate the contributions of Benjamin Lahey. Not only is this thesis inspired by his work on general psychopathology, but he was also a great help in implementing and interpreting the general psychopathology models discussed here. I also wish to thank Jason Clark and Sven Walter for sparking my interest in psychiatric genetics during the course evolutionary psychiatry.

Beyond direct colleagues, many others helped with this thesis as well. I want to thank Paul Zhutovsky for being my friend for so many years and always somehow ending up not too far away. Paul teaching me math and neuroinformatics and our discussions on many scientific topics helped with this thesis, but his support as friend meant a lot to me as well. The list of important friends would not be complete without mentioning also Stas Scherba and Lea Penning, who I could always count on joining crazy projects and who always were there to help and support me. I also want to thank my host family during my high school exchange in the US, the Ferrell family. Given how crucial it is to be able to communicate in English in science, them giving me an opportunity to practice English in the US was a large help in writing this thesis. Beyond the English skills, their introduction to a different culture and philosophies helped me think how we acquire and how we evaluate knowledge, which influenced my scientific thought processes.

My girlfriend Yu-Chin Her had a direct and indirect impact on this thesis as well. Her curiosity and directness always forced me to critically evaluate my work inside and outside academia and as a result improve. At the same time the happiness she brought into the second half of my PhD time gave me strength to complete this thesis. I also thank Yu-Chin for drawing the painting found in the cover. Finally, the thesis would not have been possible without my family: my father, Jörg Neumann, my mother, Nadia Neumann, and my sister, Julia Neumann. While I was never very good at showing it, I

am happy that I could grow up with my amazing sister. My father inspired me to be curious about the world and has always supported my scientific endeavors. Thanks to him scientific thinking and appreciation of technology was a natural part of my upbringing and was the foundation for my work as PhD student. Perhaps thanks to my mom, who always had theories about why humans or animals behave a certain way, I also developed an interest in psychology. My mom has always shown me the maximum amount of love possible, which I appreciate greatly.





