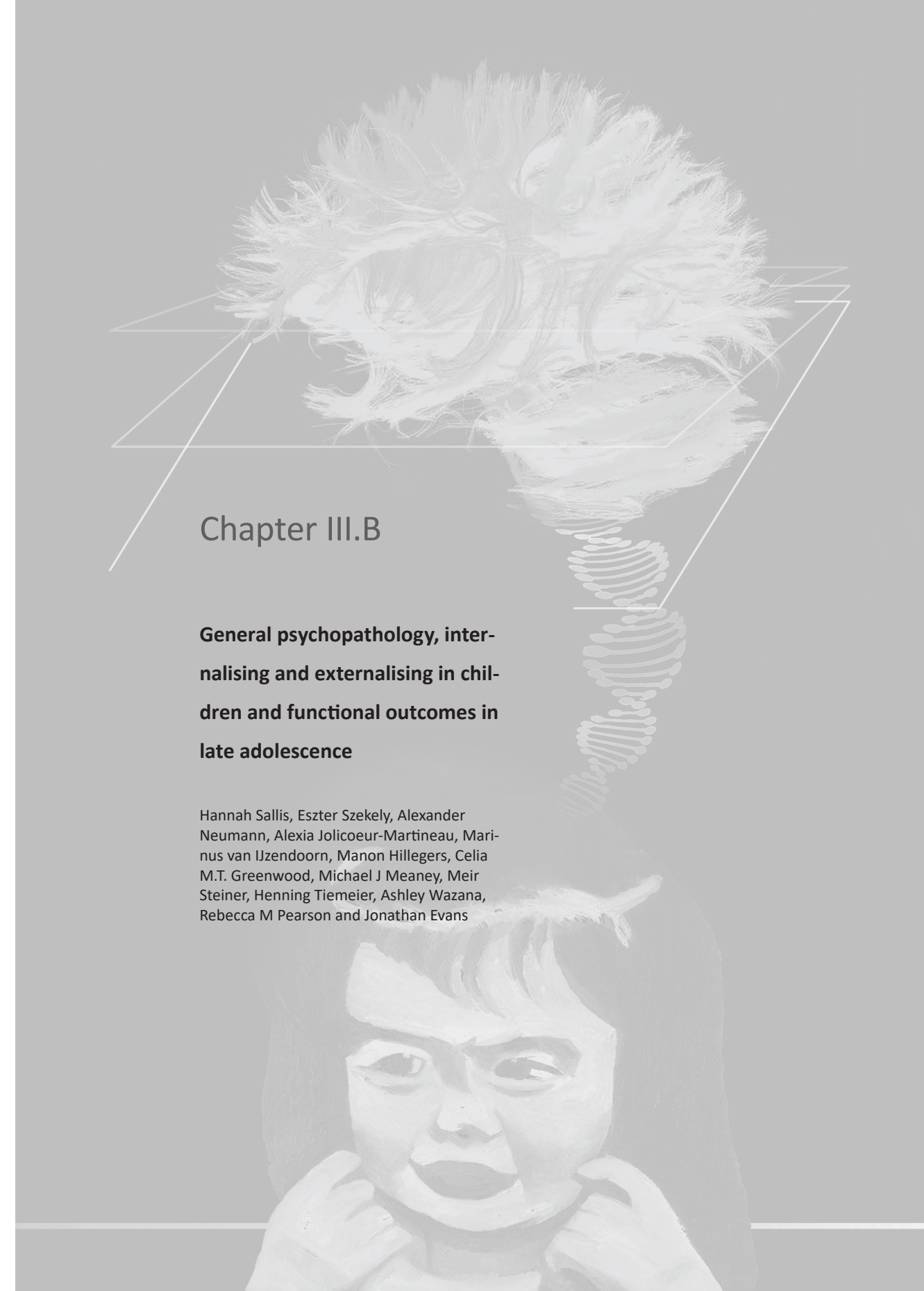

Chapter III.B

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,[1,2] the total number of disorders as well as the number of diagnoses received by each individual is rising, both for children and adults.[3] Almost half of individuals who meet diagnostic criteria for one disorder also meet diagnostic criteria for another.[4] 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.[5] In addition, the stability of these categories over time is unclear.[6–8] 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.[9] 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.[10–13] 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) ,[14,15] the Generation Rotterdam (Generation R) Study,[16,17] the Maternal Adversity, Vulnerability and Neurodevelopment (MAVAN) project,[18] and the Growing Up in Singapore Towards healthy Outcomes (GUSTO) study[19]. 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.[13,20] 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.[14–16,18,19] 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
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 & external-ising	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
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

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 (β=0.117, p=0.001), experiencing decreased psychological wellbeing at age 21 (β=-0.062, p=0.001), failing mathematics (β=-0.235, p<0.001) or English GCSE at age 16 (β =-0.260, p<0.001). Unexpectedly there was an association between GPF and reduced risk of problem drinking (β=-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 (β=0.085, p=0.030) and anxiety (β=0.184, p<0.001), decreased wellbeing (β=-0.089, p<0.001), and failure at mathematics GCSE (β=-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

			INT/EXT model (no GPF)		Bifactor model (INT, EXT, GPF)	
	Factor	**N**	**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 (β=-0.080, p=0.010) and better performance at both mathematics (β=0.050, p=0.055), and English GCSE (β=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 wellbeing 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.[21] 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.[22,23]

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_23ei4C&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, 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

18. O'Donnell KA, Gaudreau H, Colalillo S, et al. The maternal adversity, vulnerability and neurodevelopment 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.[1,2] 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.[3,4] 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.[5] 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.7 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 he 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 interview10 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 reliability11as 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.6 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 CRPS-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.16

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

	ALSPAC		Gen R		MAVAN	
Rater	**Measure (Age)**	**Subscale**	**Measure (Age)**	**Subscale**	**Measure (Age)**	**Subscale**
Parent	**DAWBA**	Depression	**CBCL**	Emotionally reactive	**CBCL**	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		
	SDQ	Emotional problems			**SDQ**	Emotional problems
	(7 years)	Peer problems			(Mother – 5 and 6 years; Father – 5 years)	Peer problems
		Conduct problems				Conduct problems
		Hyperactivity				Hyperactivity
		Prosocial				Prosocial
			CPRS-R	ADHD inattentive	**CPRS-R**	ADHD inattentive
			(8 years)	ADHD hyperactive impulsive	(Mother and father – 5 and 6 years)	ADHD hyperactive impulsive
				ODD		ODD
			SRS	Social cognition	**PAPA**	Separation anxiety

Table S1: (Continued)

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

Table S1: (Continued)

	ALSPAC		Gen R		MAVAN	
Rater	**Measure (Age)**	**Subscale**	**Measure (Age)**	**Subscale**	**Measure (Age)**	**Subscale**
	(7 years)	Attention s ymptoms score				
		Burden of attention/ activity				
		Awkward behaviours score				
		Troublesome behaviours				
		Burden of troublesome behaviours				
Field worker	**Field worker observations**	Cooperative				
	(7 years)	Fidget				
		Active				
		Attention				
		Responsive				
Child			**BPI**	Depression	**Dominic**	Depression
			(6 years)	Separation anxiety	(6 years)	Separation anxiety
				Overanxious		Overanxious
				Oppositional defiant		Oppositional defiant
				Overt hostility		Conduct disorder
				Conduct problems		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; [3] 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

Internalising factor			Externalising factor			General factor (GPF)
Rater	**Measure**	**Scale**	**Rater**	**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			Aggressive behaviour	
		Withdrawn		PAPA	ADHD	
		Somatic problems			Oppositional defiant disorder	
	PAPA	Social phobia			Conduct disorder	
		Overanxious	Mother and father	SDQ	Conduct problems	
		Panic			Hyperactivity	
		Depression			Peer problems	
Mother and father	SDQ	Emotional problems		Conner's PRS-R	Oppositional problems	
Child	Dominic	Separation anxiety			Inattention/cognitive problems	
		Overanxious			Hyperactivity	
		Simple phobias	Child	Dominic	Oppositional defiant disorder	
		Depression			Conduct disorder	

Table S5: 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
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

			No GPF (unadjusted)		Bifactor model (unadjusted)	
	Factor	**N**	**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, 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
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 neurodevelopment 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 Respponsiveness 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.