Supplementary Material

Generating polygenic risk scores

In total 9912 ALSPAC children were genotyped using the Illumina HumanHap500-quad genotyping array. Individuals were excluded on the basis of gender mismatches; minimal or excessive heterozygosity, disproportionate levels of individual missingness (>3%), insufficient sample replication (IBD <0.8), non-European ancestry (assessed by multidimensional scaling analysis and compared with Hapmap II) and cryptic relatedness (IBD > 0.1). SNPs were excluded based on minor allele frequency (<1%), call rate (<95%) or evidence for violations of Hardy-Weinberg equilibrium (P < 5E-7). Imputation was conducted by the ALSPAC team using Impute V2.2.2 against the 1000 genomes reference panel (Phase 1, Version 3: all polymorphic SNPs excluding singletons), using all 2186 reference haplotypes (including non-Europeans). SNPs were subsequently filtered based on minor allele frequency (<1%) and imputation quality (INFO<0.8). Following quality control and limiting individuals to one child per family, genetic data were available for N=7975.

Genome-wide association study (GWAS) summary statistics used to generate PRS were filtered to remove SNPs that were palindromic, insertions/deletions, non-autosomal, INFO score <0.8, missing in N>1 study and duplicates (https://github.com/ricanney). Depression results for 23andme (75,607 cases and 231,747 controls) (1) and the other samples included in the latest depression GWAS (2) (PGC29, deCODE, Generation Scotland, GERA, iPSYCH, and UK Biobank) were meta-analyzed in METAL. PRS were generated in ALSPAC using PRSice (3); SNPs were clumped with an R^2 threshold of 0.1 and a distance threshold of 1000kb and excluding the extended major histocompatibility complex (MHC; chromosome 6: 26-33Mb) due to the high linkage disequilibrium (LD) within this region.

Selecting the number of trajectories

To select the number of classes for the two growth mixture models (GMMs), we initially modelled a single k-class solution, modelling subsequent k+1 solutions until the optimum solution was reached. Residual variances were fixed across measurement points. Each model was run with 5000 random starting values and 500 optimizations (STARTS = 5000 500 in Mplus) (4) and included both linear and quadratic change. Fit statistics are presented in Table S2. Both the intercept and quadratic variance was fixed to zero because for models with more than two classes these explained the variation in the intercept and quadratic (i.e. was close to zero). Model fit significantly improved, as indicated by the fall in loglikelihood value, sample size

adjusted Bayesian information criterion as well as the Bootstrapped Likelihood Ratio Test, from the one to six class solutions. However, the six class two small classes (≤2%). Current guidance (5) suggests that if fit indices are similar (size adjusted Bayesian information criterion being the preferred index), unless there are strong theoretical reasons for preferring a particular solution, the more parsimonious solution i.e. with fewer classes is preferred. Because this is the first study to our knowledge to investigate irritability trajectories across childhood and adolescence, there is no clear theoretical guidance on how many or what shaped trajectories are to be expected. The five class solution was therefore selected and this also showed high classification accuracy (entropy = 0.94).

Inverse probability weighting

Inverse probability weighting (IPW) was used to assess the impact of missing genetic data and has been recommended over alternative methods such as multiple imputation in situations where whole blocks of data are missing for a large proportion of individuals (6). As described elsewhere (7), weights were derived from a logistic regression analysis of missing genetic data for those in the 'core' ALSPAC sample (N = 7495/13793) for a set of measures assessed in pregnancy: child gender, maternal age and child birth weight. The analyses were rerun using IPW to address any potential bias caused by only a subsample having genetic data and revealed a similar pattern of results (see Table S5).

References

1. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat Genet. 2016;48(9):1031-6.
2. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50(5):668-81.
3. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. Bioinformatics. 2015;31(9):1466-8.
4. Muthén LK, Muthén BO. Mplus User's Guide. Seventh ed. Los Angeles, CA: Muthén & Muthén; 1998-2012.
5. Wickrama KK, Lee TK, O'Neal CW, Lorenz FO. Higher-order growth curves and mixture modeling with Mplus: A practical guide: Routledge; 2016.
6. Seaman SR, White IR, Copas AJ, Li L. Combining multiple imputation and inverse-probability weighting. Biometrics. 2012;68(1):129-37.
7. Riglin L, Thapar AK, Leppert B, Martin J, Richards A, Anney R, et al. The contribution of psychiatric risk alleles to a general liability to psychopathology in early life. bioRxiv. 2018;doi:10.1101/409540.

Table S1. Irritability measure descriptive statistics

		Cumulative frequency (%)							Average	
		0	0-1	0-2	0-3	0-4	0-5	0-6	Mean	(SD)
Age 7 years 7 months	(N=6975)	76.36	87.17	93.72	97.12	98.39	99.20	100	0.48	(1.06)
Age 10 years 8 months	(N=7180)	75.84	86.31	93.30	96.85	97.97	98.86	100	0.51	(1.12)
Age 13 years 10 months	(N=6646)	78.33	87.30	93.73	97.14	98.09	98.92	100	0.46	(1.09)
Age 15 years 6 months	(N=4415)	76.81	86.14	92.46	96.44	97.40	98.53	100	0.52	(1.18)

Table S2. Model fit indices for irritability growth mixture models

	LL	Free parameters	BIC	ssaBIC	Smallest class	Entropy	VLMR-LRT p value	BLRT p value
1 class	-37494.52	5	75033.93	75018.04	100%			
2 classes	-33448.78	9	66978.35	66949.75	8%	0.96	<0.0001	<0.0001
3 classes	-32326.50	13	64769.70	64728.39	6%	0.96	0.0599	<0.0001
4 classes	-30994.71	17	62142.04	62088.02	3%	0.95	0.0009	<0.0001
5 classes*	-30023.11	21	60234.75	60168.02	2%	0.94	0.0001	<0.0001
6 classes**	-29399.46	24	59014.39	58938.12	2%	0.94	0.0004***	<0.0001***

LL=Loglikelihood; BIC= Bayesian Information Criteria; ssa= sample size adjusted; VLMR-LRT=Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest; BLRT=Bootstrapped Likelihood Ratio Rest. *Final model. **Slope variance also fixed to zero variance because the slope variance was close to zero (i.e. classes explained the variation in the slope). ***Compared to a 5 class model where slope variance was also fixed to zero.

Table S3. Multivariate associations between genetic risk and trajectories

	ADHD PRS			Depression PRS		
	OR	95% CI	p	OR	95% CI	p
Childhood-onset irritability						
Decreasing	1.05	(0.91, 1.20)	0.524	1.11	(0.96, 1.28)	0.147
Late-childhood limited	1.15	(1.00, 1.33)	0.053	0.92	(0.81, 1.04)	0.173
High-persistent	1.32	(1.09, 1.60)	0.005	0.97	(0.79, 1.17)	0.726
Adolescent-onset irritability						
Increasing	1.26	(1.09, 1.45)	0.002	1.17	(1.01, 1.34)	0.032

Low trajectory as reference. ADHD=attention-deficit/hyperactivity disorder, PRS = polygenic risk score

Table S4. Associations between genetic risk and trajectories using IPW

	Original result			Core sample			Core sample: IPW		
ADHD PRS	OR	95% CI	p	OR	95% CI	p	OR	95% CI	p
Decreasing	1.06	(0.92, 1.22)	0.425	1.09	(0.95-1.26)	0.221	1.10	(0.95-1.28)	0.201
Late-childhood limited	1.14	(0.99, 1.32)	0.071	1.14	(0.98-1.32)	0.079	1.13	(0.98-1.31)	0.093
High-persistent	1.31	(1.09, 1.58)	0.005	1.30	(1.08-1.57)	0.006	1.29	(1.07-1.56)	0.007
Increasing	1.28	(1.11, 1.48)	0.001	1.28	(1.11-1.49)	0.001	1.31	(1.13-1.53)	<0.001
	Original result			Core sample			Core sample: IPW		
Depression PRS	OR	95% CI	p	OR	95% CI	p	OR	95% CI	p
Decreasing	1.12	(0.97, 1.29)	0.126	1.14	(0.99-1.31)	0.075	1.16	(1.00-1.34)	0.043
Late-childhood limited	0.93	(0.82, 1.06)	0.266	0.93	(0.82-1.06)	0.298	0.94	(0.83-1.07)	0.382
High-persistent	1.00	(0.82, 1.21)	0.992	1.03	(0.85-1.24)	0.763	1.03	(0.85-1.25)	0.732
Increasing	1.20	(1.05, 1.38)	0.009	1.21	(1.051.40)	0.009	1.24	(1.07-1.44)	0.004

Low trajectory as reference. IPW= inverse probability weighting. ADHD=attention-deficit/hyperactivity disorder, PRS = polygenic risk score

Table S5. Association between ADHD and depression genetic risk scores and irritability trajectories, controlling for gender

	ADHD PRS			Depression PRS		
	OR	95% CI	p	OR	95% CI	p
Childhood-onset irritability						
Decreasing	1.06	(0.92, 1.22)	0.401	1.14	(0.99, 1.32)	0.121
Late-childhood limited	1.14	(0.99, 1.32)	0.068	0.93	(0.82, 1.06)	0.281
High-persistent	1.32	(1.09, 1.59)	0.004	1.01	(0.83, 1.22)	0.958
Adolescent-onset irritability						
Increasing	1.28	(1.11, 1.48)	0.001	1.20	(1.05, 1.38)	0.009

Low trajectory as reference. ADHD=attention-deficit/hyperactivity disorder, PRS = polygenic risk score

Table S6. Association between childhood ADHD and adolescent depression diagnoses and irritability trajectories, controlling for gender

	Childhood ADHD			Adolescent depression		
	OR	95% CI	p	OR	95% CI	p
Childhood-onset irritability						
Decreasing	29.70	(14.81-59.55)	<0.001	2.41	(0.92-6.34)	0.074
Late-childhood limited	19.13	(9.21-39.73)	<0.001	2.01	(0.66-6.17)	0.220
High-persistent	101.09	(53.05-192.64)	<0.001	7.68	(3.27-18.00)	<0.001
Adolescent-onset irritability						
Increasing	6.40	(2.05-20.04)	0.001	4.73	(2.26-9.90)	<0.001

Low trajectory as reference. ADHD=attention-deficit/hyperactivity disorder. OR for childhood ADHD = likelihood of being in irritability trajectory class for ADHD diagnosis vs. no ADHD diagnosis; gender predicting irritability trajectory class. OR for adolescent depression = likelihood of having adolescent depression for given irritability trajectory class vs low irritability trajectory class; gender predicting depression.

Table S7. Association between ADHD and depression genetic risk scores and irritability trajectories, excluding diagnoses

		Full sample (N=5559)			Excluding childhood ADHD (N=4892)	
ADHD PRS	OR	95% CI	p	OR	95% CI	p
Decreasing	1.06	(0.92, 1.22)	0.425	1.08	(0.92-1.25)	0.324
Late-childhood limited	1.14	(0.99, 1.32)	0.071	1.11	(0.95-1.29)	0.208
High-persistent	1.31	(1.09, 1.58)	0.005	1.31	(1.04-1.64)	0.021
Increasing	1.28	(1.11, 1.48)	0.001	1.21	(1.03-1.42)	0.021
		Full sample (N=5559)			Excluding adolescent depression (N=3718)	
Depression PRS	OR	95% CI	p	OR	95% CI	p
Decreasing	1.12	(0.97, 1.29)	0.126	1.13	(0.95-1.34)	0.160
Late-childhood limited	0.93	(0.82, 1.06)	0.266	0.88	(0.76-1.02)	0.089
High-persistent	1.00	(0.82, 1.21)	0.992	1.06	(0.81-1.40)	0.659
Increasing	1.20	(1.05, 1.38)	0.009	1.16	(0.97-1.39)	0.109

Low trajectory as reference. ADHD=attention-deficit/hyperactivity disorder, PRS = polygenic risk score

Figure S1. Participation rates

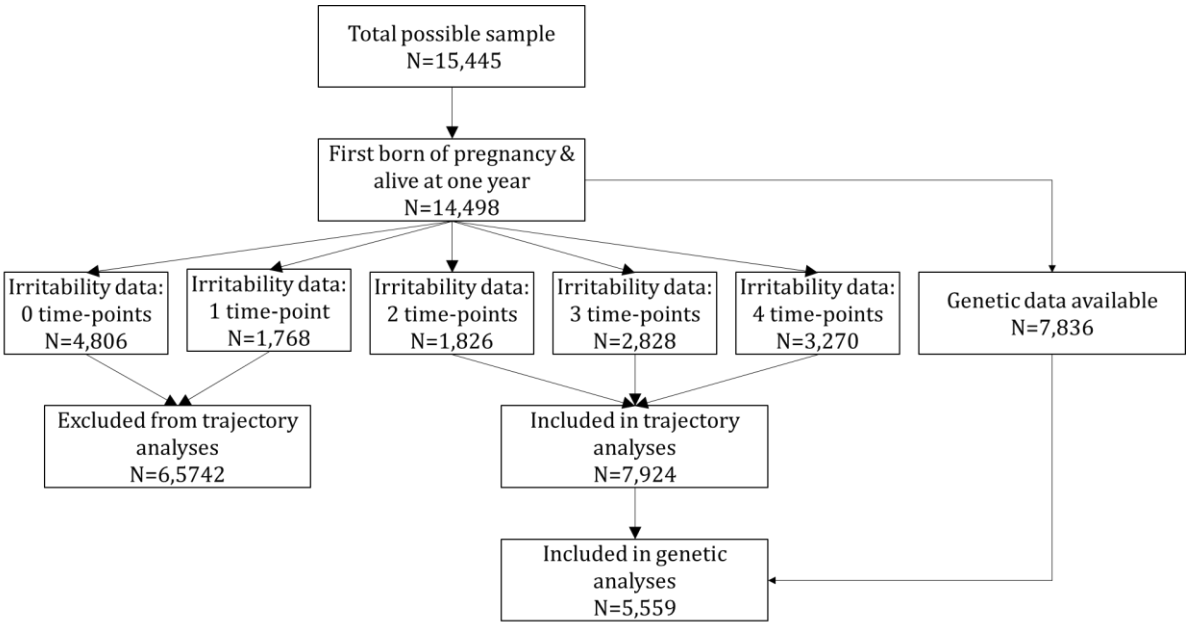
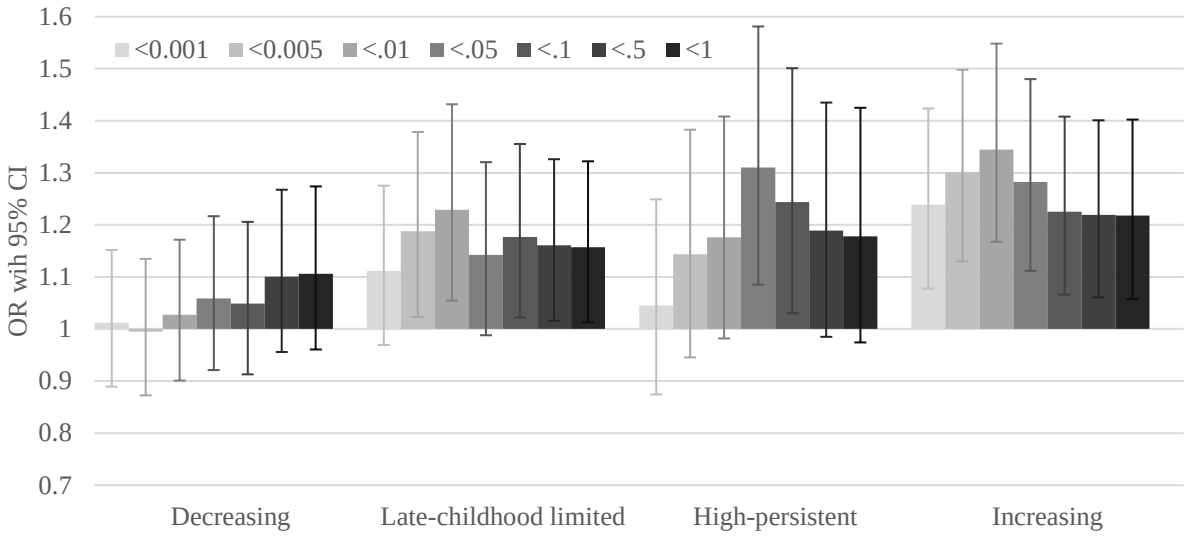


Figure S2. Polygenic risk score associations using a range of p-value thresholds from the discovery sample

a) Attention-deficit/hyperactivity disorder polygenic risk scores



b) Depression polygenic risk scores

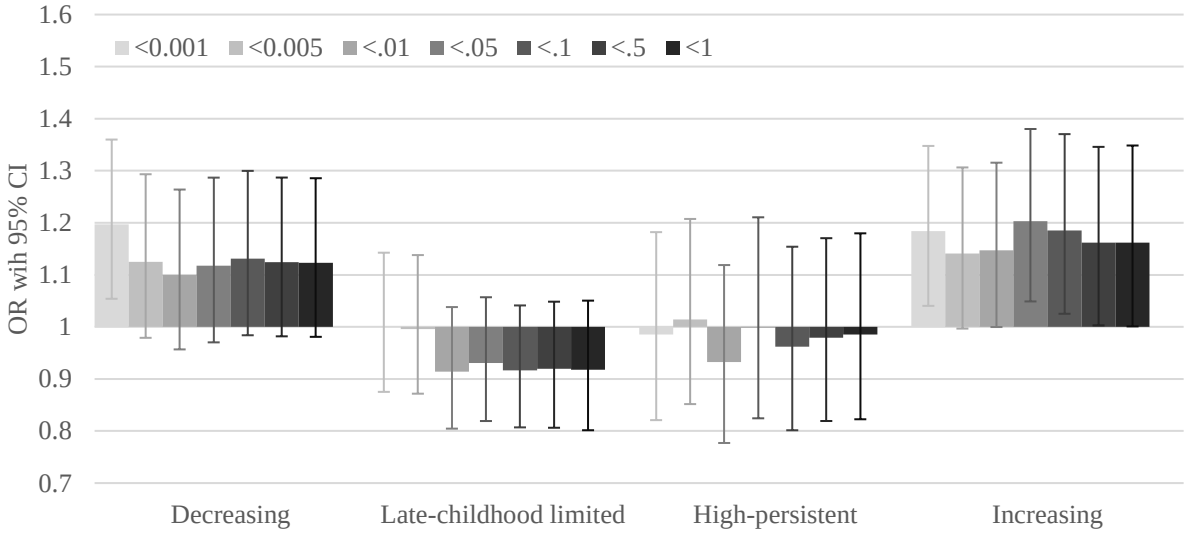
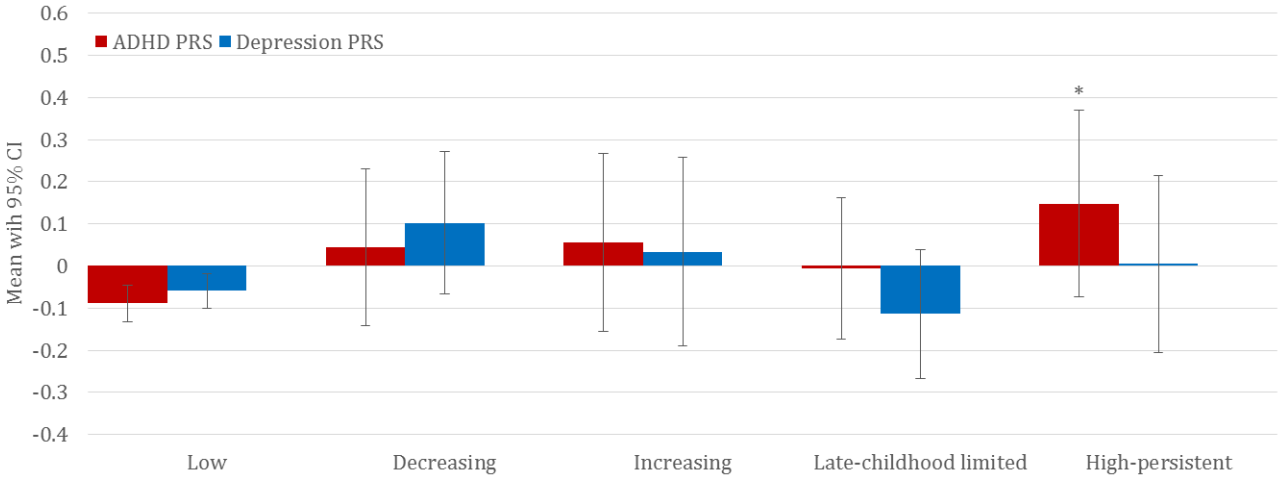


Figure S3. Mean ADHD and depression genetic risk score by irritability trajectories, with 95% confidence intervals

a) boys



b) girls

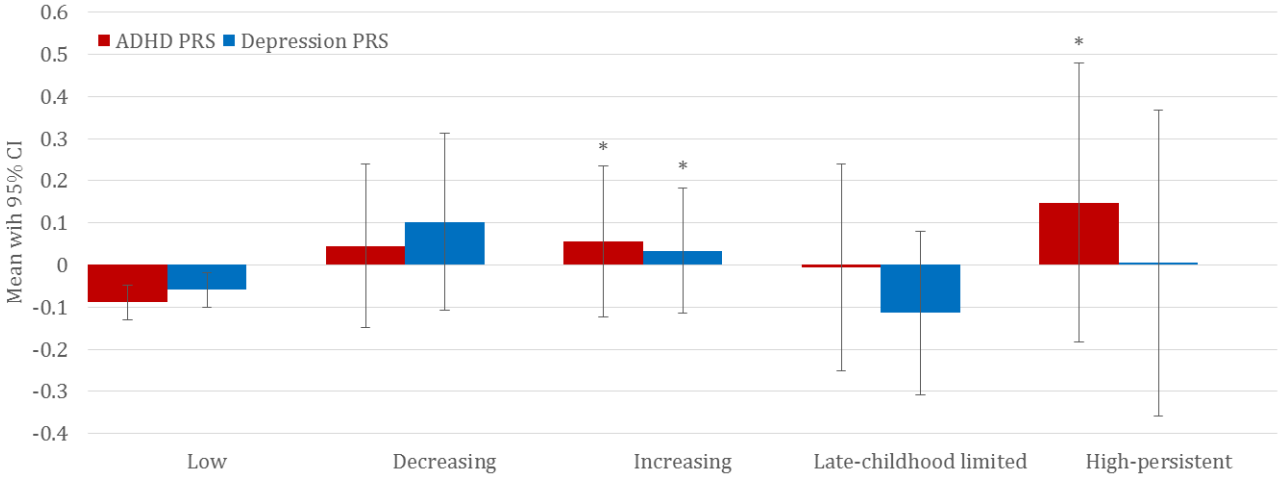
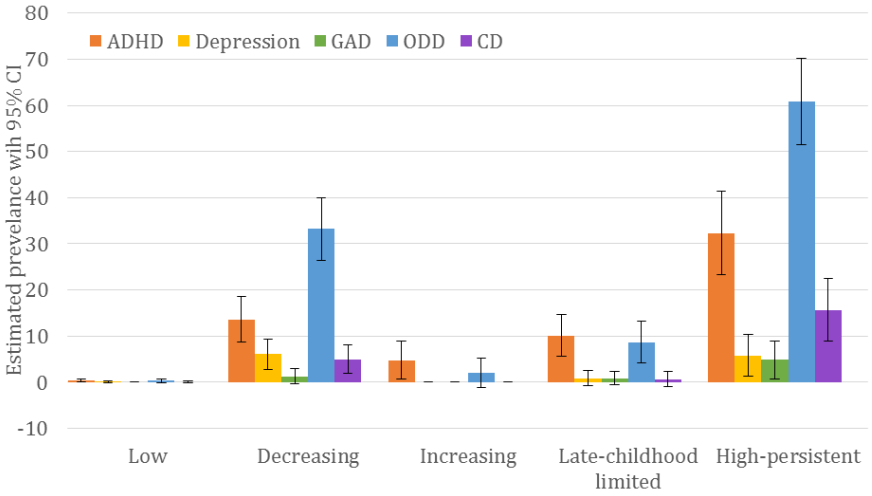
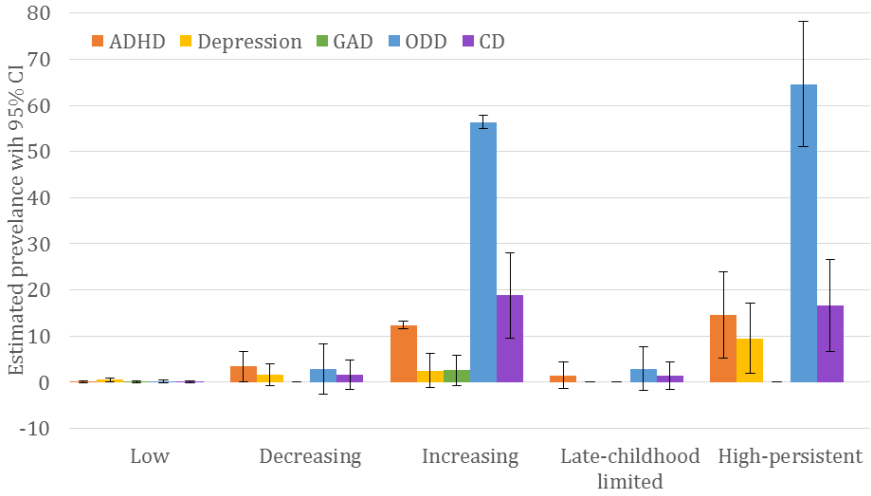


Figure S4. Estimated prevalence of diagnoses by irritability trajectories, with 95% confidence intervals

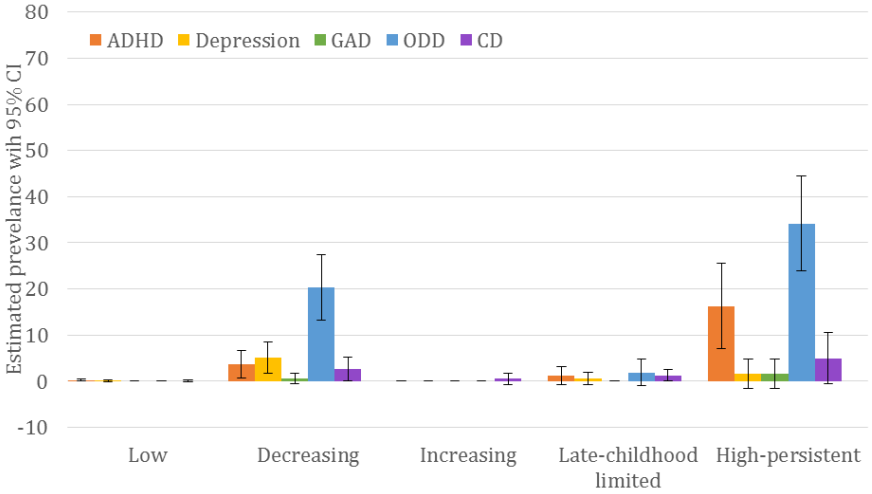
a) Boys: childhood



b) Boys: adolescence



c) Girls: childhood



d) Girls: adolescence