
Article:
Mahedy, L., Suddell, S., Skirrow, C. et al. (6 more authors) (2021) Alcohol use and cognitive functioning in young adults : improving causal inference. Addiction, 116 (2). pp. 292-302. ISSN 0965-2140

https://doi.org/10.1111/add.15100

Supplementary Material

METHODS

Cognitive functioning measures

Working memory. Participants continuously monitored a series of numbers presented on a computer screen and pressed '1' if the number was the same as the number presented *N* numbers ago, or '2' if it was not. Stimuli were numbers 0–9, presented in black on white background with a random spatial jitter of 180 pixels in y-axis and 200 pixels in x-axis. Each target was presented for 500 ms, followed by a 3,000 ms response window. The practice block consisted of 12 trials containing two targets. The experimental block consisted of 48 trials, containing 8 targets, where the target was the number that was identical to the one presented 2 trials back. Three outcomes were examined for the *N*-back task (i) number of hits, or the percentage of matching numbers correctly identified as matches, (ii) false alarms, or the percentage of non-matching numbers incorrectly identified as matches, and (iii) discriminability index, *d*', which is a signal-detection metric that takes into account both hits and false alarms to derive an overall estimate of signal-detection ability (1). *d*' was calculated using the Stata syntax adapted from (2). High scores on *number of hits* indicated more accurate identification, while high scores on *false alarms* indicated less accurate identification. High scores on *d*', therefore, indicated a greater ability to distinguish signal from noise.

$$d' = invnorm(\text{hits}) - invnorm(\text{false alarms})$$

Although there is some debate in the literature surrounding the construct validity of performance on the *N*-back task as an indicator of working memory ability, it has been argued that using *N*-back performance indices from a signal-detection framework (i.e. d') may reveal clearer insights about its validity as a measure of working memory performance (3–5).

Response inhibition. Participants were asked to sit in front of a computer monitor and their two index fingers were placed in two stimulus boxes, one labelled X and one labelled O. Two types of trials were performed, primary trials and stop signal trials. In the primary trials, participants were asked to fixate on a plus sign (+) in the centre of the computer screen. An X or O was presented on the screen and the participant had to press the corresponding button as quickly as possible. The second part is identical to the first but a beep is heard (stop signal), randomly after the X and O appears on 25% of the trials. The beep sounded on random trials at 150ms before the child's mean reaction time to the primary trials. Participants were told to not press a response button when the beep was sounded, and to wait for the next trial to begin. If the beep was not heard the participant was asked to press the corresponding key according to what was presented on screen. When the beep was sounded, the participant was to refrain from pressing the response button.

32 practice trials were presented. The task consisted of 256 trials, comprised of 4 blocks of 64 trials. Each block of 64 trials consists of 4 sub-blocks of 16 trials. Each sub-block consists of 12 trials without a stop-signal and 4 trials with a stop-signal. Mean response times were calculated. Five metrics were examined for the stop signal task: (i) an estimate of stop signal reaction (SSRT) was calculated and used as the primary outcome as it is a reliable measure of inhibitory control, with shorter SSRT's indicating slower inhibition;

secondary outcomes include: (ii) 'go' reaction time; (iii) 'stop' reaction time; (iv) 'go' accuracy; and (v) 'stop' accuracy.

$$SSRT_{med} = \text{Go Reaction Time}_{med} - \text{Stop Signal Delay}_{med}$$

Stop Signal $Delay_{med}$ (SSD) was calculated for each session using a weighted least squares linear regression to predict SSD based on the probability of responding given a stop-signal. This was then used to estimate the SSD where the probability of the participant failing to inhibit was 50%.

Emotion recognition. Prototypical composite images of the six basic facial expressions of emotion were generated from 12 individual male and female faces showing each of the six expressions. The 12 original images were each delineated with 172 feature points, which allowed both shape and colour information to be averaged across the faces to generate 'average' anger, sadness, surprise, disgust, fear, happiness, using established techniques. An overall emotional prototype face was then generated by averaging the exemplars for each emotional expression. Facial images showing a specific emotion were displayed on the screen one at a time. Images were presented for 200 ms, followed by a backwards mask (white noise) of 250 ms. Participants were required to select the descriptor that best described the emotion that was present in the face, using the computer mouse. Emotion intensity is varied across 8 stimuli within each emotion on a scale from the most prototypical emotion to an almost neutral emotion. Each individual stimuli is presented twice, giving a total of 96 trials. The task was delivered using E-Prime Professional v. 2.0 software (6).

For each of the specific emotions an unbiased hit rate was derived and used as the secondary outcome. This is based on work by Wagner (7) who proposed an alternative score, the "unbiased hit rate" (H_u), designed to account for response biases. H_u for each participant is calculated as the squared frequency of correct responses for a target emotion divided by the product of the number of stimuli representing this emotion and the overall frequency of this emotion category being chosen. H_u has a range of zero to one, one indicating that all stimuli of an emotion have been correctly identified and the respective emotion has never been falsely chosen for a different emotion. Results from the secondary analyses are presented in Tables S8a-d.

Potential confounders

Confounders included: income (quintiles), maternal education (<O level: indicating no qualification; O level: indicating completion of school examinations at age 16; and >O level: indicating completion of college or university education at or after age 18), socioeconomic position (SEP, grouped into four categories: (a) unskilled or semiskilled manual; (b) skilled manual or non-manual; (c) managerial and technical and (d) professional), housing tenure (mortgaged, subsidised renting and private renting), sex, and maternal smoking during first trimester in pregnancy (yes/no).

A computerized version of the Counting Span task (8) was included at approximately 11 years (*M*=10 years 8 months, SD=3 months) to assess working memory performance during a clinic visit. A span score was based on the number of correctly recalled sets (maximum score of 5 in increments of 0.5). Since adolescents who have experienced head

injury perform poorly in working memory tasks compared with age-matched peers (9), we covaried for head injury/unconsciousness before the age of 11, *n*=113 (3.4%).

A modal class variable defined three patterns in the data based on longitudinal latent class analysis (LLCA): no use (*n*=4,533, 85.5%); cigarettes only (*n*=412, 7.8%); and cigarettes and/or cannabis (*n*=355, 6.7%). Briefly, this approach assigns individuals to their most likely class based on their probability of belonging. See Jones et al. (10) for a detailed description of the latent classes.

Missing data

Missing data on the binge drinking measures were dealt with using full information maximum likelihood. SES confounders assessed largely in pregnancy had minimal missing data (e.g., parental social class had the most amount of missing data: 817/9,600 (8.5%), while the cognitive measures assessed up to age 11 years and substance use assessed at age 16.5 years had moderate missing data 2,077/6,888 (30.1%). Given that the BCH method uses listwise deletion for the outcome measures, *n*=2,086 participants had complete information on outcome and confounder data and at least one measure of binge drinking.

Inverse probability weighting

Weights were derived from logistic regression models using variables associated with nonresponse, including maternal age, grandmother having a history of severe depression, maternal alcohol use in pregnancy, financial problems, maternal cannabis use and financial problems. We weighted the included respondents by the inverse of the probability of attending and used the Hosmer-Lemeshow test to assess model fit.

Model fit LLCA of binge drinking

Criteria for best fit included i) information-theoretic methods with lower values indicating better fit to the data i.e., sample size adjusted Bayesian Information Criterion (SSABIC) (11), Akaike's Information Criteria (AIC) (12), Bayesian Information Criterion (BIC) (13); and ii) likelihood ratio statistical test methods comparing the model with K classes to a model with K-1 classes i.e., Lo-Mendell-Rubin likelihood ratio test (LRT; Lo, Mendell, & Rubin, 2001) (14), bootstrap likelihood ratio test (BLRT) (15). We repeated the estimation procedure while varying the amount of missing data.

Patterns of binge drinking – class validation

Table S4 demonstrates that the pattern of binge drinking classes and later alcohol use was as expected. There was a stepped increase for all binge drinking classes with increasing risk of DSM-V alcohol dependence at age 24 years compared to the low binge drinking class: 'adolescent limited' (*b*=-0.24; 95% CI=-0.77, 0.28); 'adult limited' (*b*=0.84; 95% CI=0.58, 1.09); and 'early onset regular binge drinking' (*b*=1.58; 95% CI=1.09, 2.07), Wald test = 84.02(3), *p*<.001.

A similar but stronger pattern of results was evident examining patterns of binge drinking and performance on the AUDIT-C at age 24: 'adolescent limited' (*b*=1.24; 95% CI=0.18, 2.30); 'adult limited' (*b*=3.72; 95% CI=3.25, 4.20); and 'early onset regular binge drinking' (*b*=5.81; 95% CI=5.01, 6.62), Wald test = 341.87(3), *p*<.001. This indicates that there is a dose response relationship between the patterns of binge drinking and scores on the AUDIT-C.

Latent growth model of binge drinking

Youth binge drinking scores across all five time points were used. A latent growth model of the five repeated measures of binge drinking was conducted to examine the association with working memory, response inhibition, and emotion recognition while controlling for potential confounding variables (*n*=3,155). The intercept factor loadings were all fixed at one and the slope factor loadings were fixed to reflect the amount of time in months between assessments with baseline at zero. A quadratic growth model with no quadratic variance/covariance terms was specified. Summary fit statistics were examined to assess model fit. On the basis of summary fit statistics, a quadratic model with zero variance and covariances for the quadratic growth factor was deemed acceptable: CFI = 0.977, TLI = 0.931, and RMSEA = 0.033. We found evidence of an association between baseline binge drinking status and better ability to recognise emotions. There was further evidence between growth of binge drinking over time and better working memory and emotion recognition performance. Results are presented in the Supplementary material (Figure S3 and Table S10).

Genetic data

ALSPAC children were genotyped using the Illumina HumanHap550 quad chip genotyping platforms by 23andme subcontracting the Wellcome Trust Sanger Institute, Cambridge, UK and the Laboratory Corporation of America, Burlington, NC, US. The resulting raw genome-wide data were subjected to standard quality control methods. Individuals were excluded on the basis of gender mismatches; minimal or excessive heterozygosity; disproportionate levels of individual missingness (>3%) and insufficient sample replication (IBD < 0.8). Population stratification was assessed by multidimensional scaling analysis and compared with HapMap II (release 22) European descent (CEU), Han Chinese, Japanese and Yoruba reference populations; all individuals with non-European ancestry were removed. SNPs with a minor allele frequency of < 1%, a call rate of < 95% or evidence for violations of Hardy-Weinberg equilibrium ($p < 5 \times 10^{-7}$) were removed. Cryptic relatedness was measured as proportion of identity by descent (IBD > 0.1). Related subjects that passed all other quality control thresholds were retained during subsequent phasing and imputation. 9,115 subjects and 500,527 SNPs passed these quality control filters. ALSPAC mothers were also genotyped following a similar procedure, details of which are reported elsewhere (16).

We combined 477,482 SNP genotypes in common between the sample of mothers and sample of children. We removed SNPs with genotype missingness above 1% due to poor quality (11,396 SNPs removed) and removed a further 321 subjects due to potential ID mismatches. This resulted in a dataset of 17,842 subjects containing 6,305 duos and 465,740 SNPs (112 were removed during liftover and 234 were out of HWE after combination). We estimated haplotypes using ShapeIT (v2.r644) which utilises relatedness during phasing. We obtained a phased version of the 1000 genomes reference panel (Phase 1, Version 3) from the Impute2 reference data repository (phased using ShapeIT v2.r644, haplotype release date Dec 2013). Imputation of the target data was performed using Impute V2.2.2 against the reference panel (all polymorphic SNPs excluding singletons), using all 2186 reference haplotypes (including non-Europeans). This resulted in 8,237 eligible ALSPAC children with available genotype data after exclusion of related subjects using cryptic relatedness measures described previously.

Genetic Analyses

Genome-wide association studies (GWAS) were conducted for each cognitive measure (working memory, emotion recognition and response inhibition) using all ALSPAC participants who completed the cognitive assessments and had available genetic data (*n* = 2,471, *n* = 2,560, and *n* = 2,446, respectively). The same cognitive outcomes as in the observational analyses were used. Linear regression was conducted using SNPtest.2.5.0 to test associations between each single nucleotide polymorphism (SNP) and each cognitive phenotype under an additive model, controlling for age, sex, and the first 10 genetic principal components (to account for population stratification). SNPs reaching $p<5x10^{-8}$ were identified as genome-wide significant. Phenotypes were quantile normalized (using SNPtest) prior to analysis. Quality control checks were conducted on the summary data. SNPs were excluded if they deviated from Hardy-Weinberg equilibrium (at $p < 5x10^{-7}$), info of < 80%, and/or a minor allele frequency of < 1%. SNPs reaching $p < 5x10^{-8}$ were considered genome-wide significant. SNPs were then clumped to ensure independence at linkage disequilibrium (LD) $r^2 = 0.001$ and a distance of 10,000 kb, using the "clump_data" command in the TwoSampleMR R package (17).

The inverse-variance weighted (IVW) approach was used as a primary analysis, with three complementary estimation methods as sensitivity analyses which each make different assumptions about the nature of horizontal pleiotropy (where the genetic variant associates with the outcome via an independent pathway to the exposure): MR Egger (18), weighted median (19), and weighted mode (20). To test the suitability of the MR Egger method, the I^2_{GX} statistic was calculated to quantify the degree of regression dilution bias due to measurement error of SNP-exposure effects (21). The mean F statistic as an indicator of instrument strength was also calculated (Table S14). Steiger filtering was conducted to confirm the direction of effect (22). A consistent effect across all of these methods would provide the most confidence that any observed effects are not due to pleiotropy.

Results

Genetic Analyses

GWAS did not identify any genome-wide significant SNPs (at the $p < 5x10^{-8}$ level) for any of the three cognitive outcomes, therefore we report the lead SNPs at a $p < 5x10^{-6}$ threshold (Tables 1-3). The full distribution of results can be seen in Manhattan plots (Figures 1-3). QQ plots, and corresponding lambda (λ) statistics, are displayed in figures 4-6. The lead SNP for working memory was rs35225200 (chromosome 4, b = -0.27, SE = 0.05, effect allele = C, $p = 7.98x10^{-7}$, intergenic variant), emotion recognition: rs72739201 (chromosome 9, b = -0.32, SE = 0.07, effect allele = C, $p = 1.63x10^{-6}$, intergenic variant), and response inhibition: rs112422339 (chromosome 4, b = 0.37, SE = 0.07 effect allele = G, $p = 4.62x10^{-7}$, intronic variant within gene MARCHF1). See Tables S11 to S15.

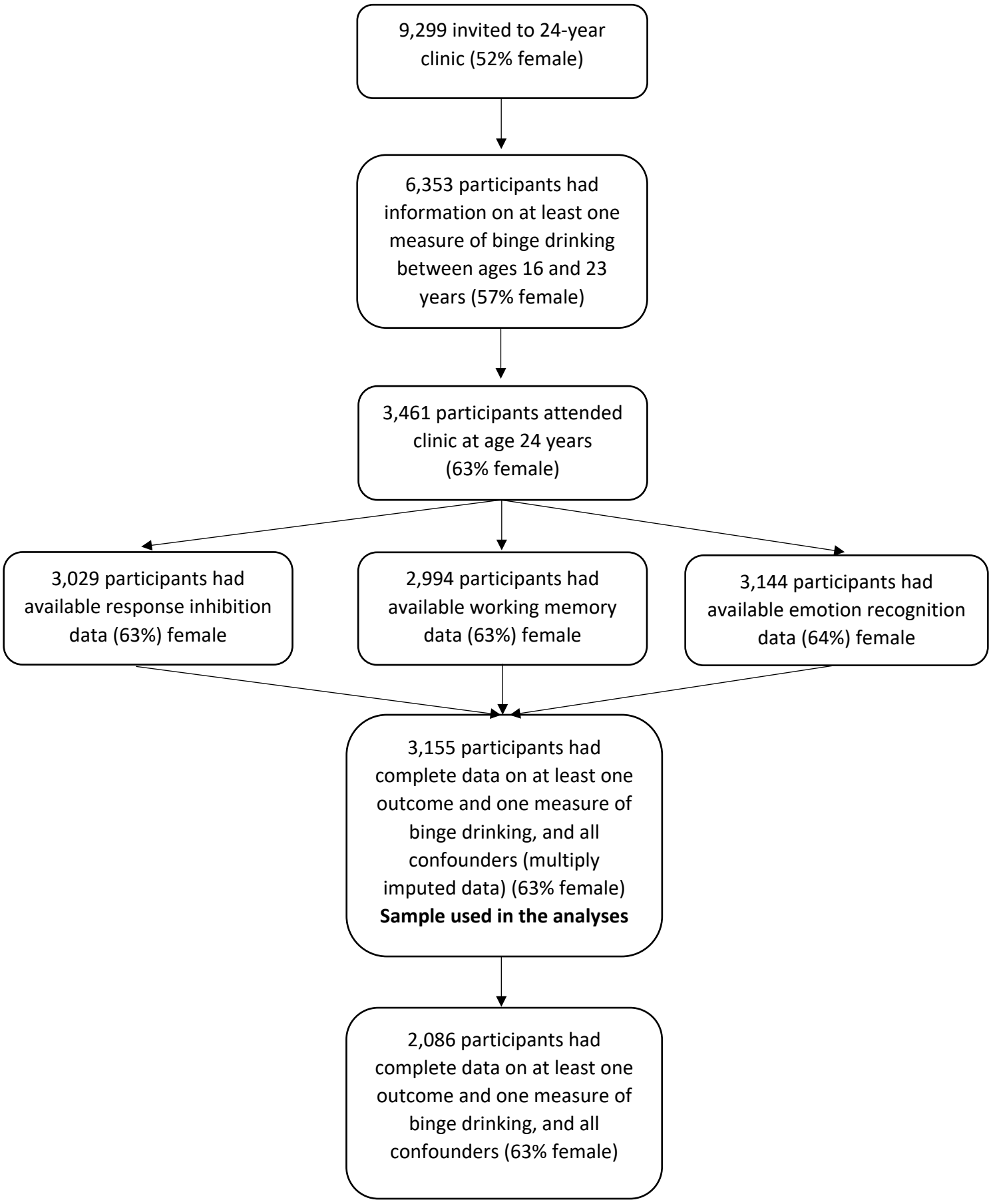

Figure S1. Sample attrition in ALSPAC

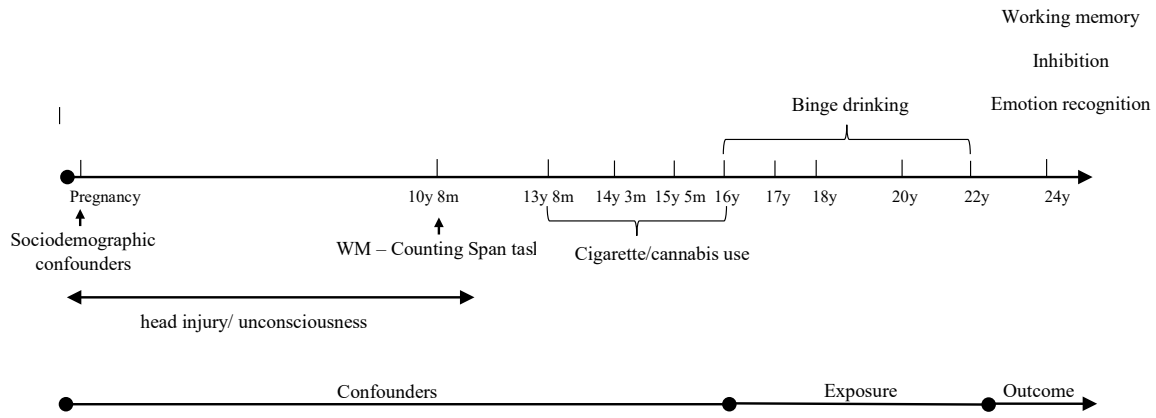

Figure S2. Timeline for data collection

Table S1. Selective attrition for cognitive functioning assessed at the age 24 clinic

	Available (*n*=3,155) *n* (%)	Not available (*n*=10,823) *n* (%)	OR (95% CI)
Gender:			
Males	1,208 (37.7)	6,009 (55.8)	0.48 (0.44, 0.52)
Income:			
Low 20%	359 (126)	1,630 (23.1)	Ref
40%	486 (17.0)	1,480 (21.0)	0.67 (0.58, 0.78)
60%	575 (20.1)	1,401 (19.8)	0.53 (0.46, 0.62)
80%	679 (23.7)	1,306 (18.5)	0.42 (0.37, 0.49)
Highest %	760 (26.6)	1,247 (17.7)	0.36 (0.31, 0.42)
Maternal education:			
<O level	1,559 (50.1)	2,826 (30.4)	Ref
O level	1,040 (33.5)	3,247 (35.0)	1.72 (1.57, 1.89)
>O level	509 (16.4)	3,214 (34.6)	3.48 (3.11, 3.90)
Social:			
iv-v	89 (3.0)	593 (7.0)	Ref
iii	929 (30.1)	3,547 (48.8)	0.57 (0.45, 0.72)
ii	1,379 (45.9)	3,420 (40.3)	0.37 (0.30, 0.47)
Professional	605 (20.2)	919 (10.8)	0.23 (0.18, 0.29)
Tenure:			
Mortgaged	2,687 (86.4)	6,853 (69.3)	Ref
Private rent	233 (7.5)	1,152 (11.6)	1.94 (1.67, 2.24)
Sub rent	190 (6.1)	1,890 (19.1)	3.90 (3.34, 4.56)
Maternal smoking:			
Yes	354 (12.0)	2,172 (23.5)	2.27 (2.00, 2.56)
Head injury:			
Yes	106 (3.4)	225 (3.4)	0.98 (0.80, 1.21)
Cigarette/cannabis:			
None	2,321 (87.0)	2,212 (84.0)	Ref
Smoking only	178 (6.7)	234 (8.9)	1.38 (1.13, 1.69)
Smoking and cannabis	168 (6.3)	187 (7.1)	1.17 (0.94, 1.45)
WM at age 11:	M (SD)	M (SD)	
Linear term	3.51 (0.83)	3.36 (0.86)	0.82 (0.77, 0.86)

Note: Maternal education: <O level indicating no qualification; O level: indicating completion of school examinations at age 16; and >O level: indicating completion of college or university education at or after age 18; SEP grouped into 4 categories: iv-v: unskilled or semiskilled manual; iii: skilled manual or nonmanual; ii: managerial and technical; and i: professional; lifetime cigarette smoking up to 16.5 years of age; lifetime cannabis use up to 16.5 years of age

Table S2. Prevalence of binge drinking at each timepoint estimated using all available data

	16	%	17	%	18	%	21	%	23	%
	5,026		4,179		3,333		4,168		3,901	
Never/occ	3,328	66.2	2,608	62.4	1,625	48.8	1,719	41.2	2,070	53.1
Monthly	1,134	22.6	1,009	24.1	944	28.3	1,241	29.8	1,083	27.8
Weekly	564	11.2	562	13.5	764	22.9	1,208	29.0	748	19.2

Note: Questionnaires assessments at ages 16, 17, 21, and 23; clinic assessment at age 18

Table S3. Comparison of model fit indices comparing 1 to 5 classes

	# param	AIC	BIC	SSABIC	Entropy	Min class	LRT	BLRT
1 class	10	38190	38258	38226	-	-	-	-
2 class	21	35654	35797	35731	0.54	44.4%	<.001	<.001
3 class	32	35313	35532	35431	0.49	23.2%	<.001	<.001
4 class	43	35102	35396	35259	0.50	17.0%	.0001	.0001
5 class	54	35062	35431	35260	0.50	9.8%	.05	<.001

Table S4. Patterns of binge drinking from 16 to 23 years and alcohol dependence and AUDIT-C at age 24 years

	Low risk	Adolescence limited	Adult limited	Early onset regular	
n=3,155	Reference	*b* (95% CI)	*b* (95% CI)	*b* (95% CI)	Wald (df) *p* value
Alcohol dependence (DSM-V)	-	-0.24 (-0.77, 0.28)	0.84 (0.58, 1.09)	1.58 (1.09, 2.07)	84.02 (3) p<.001
AUDIT-C	-	1.24 (0.18, 2.30)	3.72 (3.25, 4.20)	5.81 (5.01, 6.62)	341.87 (3) p<.001

Table S5. Factors associated with binge drinking latent class membership

N=3,755		Adolescent limited	Adult limited	Early onset regular	*p* value
	N (%)	*b* (95% CI)	*b* (95% CI)	*b* (95% CI)	
Gender:					
Males	2,973 (43.2)	-0.14 (-.57, .28)	-0.30 (-.62, .02)	-1.10 (-1.42, -.94)	<.001
Income:					
Highest	1,420 (23.5)	ref	ref	ref	<.001
80%	1,340 (22.2)	-0.51 (-1.15, .13)	-0.28 (-.71, .15)	-0.38 (-.80, .03)	
60%	1,221 (20.2)	-0.83 (-1.51, -.14)	-0.61 (-1.10, -.12)	-0.37 (-.80, .05)	
40%	1,120 (18.6)	-0.55 (-1.21, .11)	-0.17 (-.70, .36)	-0.67 (-1,19, -.11)	
Lowest 20%	935 (15.5)	-0.93 (-1.66, -.18)	-0.67 (-1.35, .02)	-1.16 (-1.83, -.48)	
Maternal education:					
>O level	1,405 (21.2)	ref	ref	ref	<.001
O level	2,264 (34.2)	-0.54 (-1.11, .03)	0.89 (.26, 1.51)	0.17 (-.33, .67)	
<O level	2,949 (44.6)	-0.29 (-.80, .23)	0.53 (-.11, 1.16)	0.09 (-.40, .58)	
Social:					
Professional	1,085 (17.2)	ref	ref	ref	<.001
ii	2,831 (44.8)	0.97 (-.18, 2.10)	-0.20 (-.60, .22)	-0.12 (-.50, .25)	
iii	2,140 (33.9)	1.02 (-.16, 2.20)	-0.47 (-.97, .04)	-0.64 (-1.11, -.16)	
iv-v	266 (4.2)	0.78 (-.96, 2.54)	0.17 (-.99, 1.33)	-0.34 (-1.35, .67)	
Tenure:					
Mortgaged	5,479 (82.3)	ref	ref	ref	<.001
Private rent	576 (8.7)	0.23 (-.84, .90)	-0.14 (-.87, .59)	0.70 (.17, 1.23)	
Sub rent	600 (9.0)	-0.12 (-.83, .58)	-1.56 (-3.09, -.02)	-0.30 (-1.04, .43)	
Maternal smoking:					
Yes	897 (14.2)	0.68 (.16, 1.20)	-0.17 (-.87, .52)	0.43 (-.06, .93)	.003
Head injury:					

Yes	222 (3.5)	0.54 (-.47, 1.55)	0.19 (-.58, .95)	0.18 (-.64, 1.00)	.18
Cigarette/cannabis:					
None	3,463 (71.7)	ref	ref	ref	<.001
Smoking only		1.44 (.69, 2.19)	-0.06 (-1.06, .93)	2.04 (1.49, 2.59)	
Smoking and cannabis		1.94 (.92, 2.95)	----	2.56 (1.88, 3.24)	
WM at age 11:	M (SD)				
Linear term	3.45 (0.85)	0.10 (-.14, .34)	0.15 (-.04, .35)	0.45 (.27, .63)	.005

Table S6. Patterns of binge drinking from 16 to 23 years and working memory measures at age 24 (high scores reflect better performance)

N=3,155 for all models	Low risk Reference group	Adolescence limited *b* (95% CI)	Adult limited *b* (95% CI)	Early onset regular *b* (95% CI)	Wald (df) p value
Unadjusted models					
Number of hits	-	-0.03 (-.08, .03)	0.04 (.01, .07)	0.03 (-.00, .06)	9.95 (3) *p*=0.02
False alarms	-	0.03 (-.03, .01)	-0.01 (-.03, .02)	-0.01 (-.03, .01)	3.01 (3) *p*=0.39
Adjusted for SES					
Number of hits	-	-0.03 (-.08, .03)	0.03 (-.00, .06)	0.02 (-.02, .05)	5.31 (3) *p*=0.15
False alarms	-	0.03 (-.02, .07)	-0.00 (-.03, .02)	-0.01 (-.03, .02)	1.47 (3) *p*=0.69
Adjusted for SES/WM/HI					
Number of hits	-	-0.03 (-.09, .02)	0.03 (-.01, .06)	0.01 (-.02, .04)	5.06 (3) *p*=0.17
False alarms	-	0.03 (-.02, .07)	-0.00 (-.03, .02)	-0.00 (-.03, .02)	1.44 (3) *p*=0.70
Fully adjusted models					
Number of hits	-	-0.03 (-.08, .03)	0.03 (-.01, .06)	0.01 (-.02, .05)	4.81 (3) *p*=0.19
False alarms	-	0.02 (-.02, .07)	-0.00 (-03, .02)	-0.01 (-.03, .02)	1.34 (3) *p*=0.72

Note. SES: socioeconomic status; WM: working memory at age ~11 years; HI: head injury/ unconsciousness up to age 11 years

Table S7. Patterns of binge drinking from 16 to 23 years and response inhibition measures at age 24 (faster times reflect better performance)

Unadjusted models	Low risk	Adolescence limited	Adult limited	Early onset regular	
N=3,155	Reference group	*b* (95% CI)	*b* (95% CI)	*b* (95% CI)	Wald (df) p value
Go reaction time	-	8.11 (-2.36, 18.58)	-2.65 (-9.32, 4.01)	-3.70 (-10.39, 3.00)	4.41 (3) *p*=0.22
Go accuracy	-	-0.01 (-.03, .01)	0.01 (-.00, .02)	-0.01 (-02, .00)	6.53 (3) *p*=0.09
Stop accuracy	-	-0.03 (-.07, .01)	0.02 (-.00, .05)	0.01 (-.02, .03)	6.59 (3) *p*=0.09
Adjusted for SES					
Go reaction time	-	8.88 (-1.77, 19.52)	-0.58 (-7.61, 6.45)	-1.44 (-8.35, 5.47)	2.93 (3) *p*=0.40
Go accuracy	-	-0.01 (-.03, .01)	0.00 (-.01, .01)	-0.01 (-.02, .00)	6.05 (3) *p*=0.11
Stop accuracy	-	-0.03 (-.07, .01)	0.02 (-.01, .04)	0.00 (-.02, .03)	4.23 (3) *p*=0.24
Adjusted for SES/WM/HI					
Go reaction time	-	9.12 (-1.58, 19.81)	-0.41 (-7.40, 6.59)	-0.58 (-7.55, 6.40)	2.94 (3) *p*=0.40
Go accuracy	-	-0.01 (-.03, .01)	0.00 (-.01, .01)	-0.01 (-.02, -.00)	8.24 (3) *p*=0.04
Stop accuracy	-	-0.03 (-.07, .01)	0.02 (-.01, .04)	-0.01 (-.03, .03)	4.60 (3) *p*=0.20
Fully adjusted model					
Go reaction time	-	9.11 (-1.90, 20.11)	-0.41 (-7.41, 6.59)	-0.58 (-8.05, 6.88)	2.79 (3) *p*=0.43
Go accuracy	-	-0.01 (-.03, .01)	0.00 (-.01, .01)	-0.01 (-.02, .00)	4.67 (3) *p*=0.20
Stop accuracy	-	-0.03 (-.07, .01)	0.02 (-.01, .04)	0.00 (-.03, .03)	3.82 (3) *p*=0.28

Note. SES: socioeconomic status; WM: working memory at age ~11 years; HI: head injury/ unconsciousness up to age 11 years

Table S8a. Patterns of binge drinking from 16 to 23 years and specific emotion sensitivity at age 24 (high scores reflect better performance)

Unadjusted models	Low risk	Adolescence limited	Adult limited	Early onset regular	
N=3,155	Reference group	*b* (95% CI)	*b* (95% CI)	*b* (95% CI)	Wald (df) p value
Anger	-	-0.03 (-.08, .02)	0.03 (-.00, .06)	0.02 (-.01, .05)	6.66 (3) *p*=0.08
Disgust	-	0.01 (-.04, .05)	0.02 (-.01, .05)	-0.00 (-.03, .03)	2.55 (3) *p*=0.47
Fear	-	-0.01 (-.07, .05)	0.04 (.00, .08)	0.03 (-.01, .07)	5.95 (3) *p*=0.11
Happy	-	-0.03 (-.07, .00)	0.01 (-.02, .03)	-0.01 (-.03, .01)	5.58 (3) *p*=0.13
Sad	-	-0.02 (-.06, .01)	0.01 (-.02, .03)	0.00 (-.02, .02)	2.34 (3) *p*=0.51
Surprise	-	-0.03 (-.07, .00)	-0.00 (-.02, .02)	0.00 (-.02, .02)	3.61 (3) *p*=0.31

Table S8b. Patterns of binge drinking from 16 to 23 years and specific emotion sensitivity at age 24 (high scores reflect better performance)

Models adjusted for SES	Low risk	Adolescence limited	Adult limited	Early onset regular	
N=3,155	Reference group	*b* (95% CI)	*b* (95% CI)	*b* (95% CI)	Wald (df) p value
Anger	-	-0.02 (-.07, .03)	0.02 (-.01, .06)	0.02 (-.02, .05)	3.80 (3) *p*=0.28
Disgust	-	0.01 (-.04, .06)	0.02 (-.02, .05)	-0.01 (-.04, .03)	1.53 (3) *p*=0.68
Fear	-	0.00 (-.06, .07)	0.03 (-.01, .07)	0.03 (-.02, .07)	2.65 (3) *p*=0.45
Happy	-	-0.03 (-.07, .00)	0.01 (-.02, .03)	-0.00 (-.03, .02)	4.53 (3) *p*=0.21
Sad	-	-0.02 (-.06, .02)	0.00 (-.02, .03)	0.00 (-.02, .03)	1.20 (3) *p*=0.76
Surprise	-	-0.03 (-.06, .01)	-0.01 (-.03, .02)	0.00 (-.02, .02)	2.29 (3) *p*=0.51

Note. SES: socioeconomic status

Table S8c. Patterns of binge drinking from 16 to 23 years and specific emotion sensitivity at age 24 (high scores reflect better performance)

Models adjusted for SES/WM/HI	Low risk	Adolescence limited	Adult limited	Early onset regular	
N=3,155	Reference group	*b* (95% CI)	*b* (95% CI)	*b* (95% CI)	Wald (df) p value
Anger	-	-0.03 (-.07, .02)	0.02 (-.01, .06)	0.01 (-.02, .04)	3.51 (3) *p*=0.32
Disgust	-	0.01 (-.04, .06)	0.02 (-.02, .04)	-0.01 (-.04, .02)	1.76 (3) *p*=0.62
Fear	-	0.00 (-.06, .06)	0.03 (-.01, .07)	0.02 (-.02, .06)	2.05 (3) *p*=0.56
Happy	-	-0.03 (-.07, .00)	0.01 (-.02, .03)	-0.01 (-.03, .02)	4.77 (3) *p*=0.19
Sad	-	-0.02 (-.06, .02)	0.00 (-.02, .03)	-0.00 (-.03, .02)	1.26 (3) *p*=0.74
Surprise	-	-0.03 (-.06, .01)	-0.01 (-.03, .02)	-0.00 (-.03, .02)	2.38 (3) *p*=0.50

Note. SES: socioeconomic status; WM: working memory at age ~11 years; HI: head injury/ unconsciousness up to age 11 years

Table S8d. Patterns of binge drinking from 16 to 23 years and specific emotion sensitivity at age 24 (high scores reflect better performance)

Fully adjusted models	Low risk	Adolescence limited	Adult limited	Early onset regular	
N=3,155	Reference group	*b* (95% CI)	*b* (95% CI)	*b* (95% CI)	Wald (df) p value
Anger	-	-0.03 (-.08, .03)	0.02 (-.01, .06)	0.01 (-.02, .04)	3.44 (3) *p*=0.33
Disgust	-	0.01 (-.04, .06)	0.02 (-.02, .05)	-0.01 (-.04, .02)	1.79 (3) *p*=0.62
Fear	-	0.01 (-.06, .07)	0.03 (-.01, .07)	0.02 (-.02, .07)	2.39 (3) *p*=0.50
Happy	-	-0.03 (-.07, .01)	0.01 (-.02, .03)	-0.00 (-.03, .02)	3.56 (3) *p*=0.31
Sad	-	-0.02 (-.05, .02)	0.00 (-.02, .03)	0.00 (-.02, .03)	0.86 (3) *p*=0.84
Surprise	-	-0.02 (-.06, .01)	-0.01 (-.03, .02)	0.00 (-.02, .02)	1.98 (3) *p*=0.58

Note. SES: socioeconomic status; WM: working memory at age ~11 years; HI: head injury/ unconsciousness up to age 11 years

Table S9. Patterns of binge drinking and working memory, response inhibition and emotion recognition - complete case analyses (fully adjusted models)

	Low risk	Early-onset monthly	Adult frequent	Early-onset frequent	
n=1,936 for all models	Reference group	*b* (95% CI)	*b* (95% CI)	*b* (95% CI)	Wald (df) *p* value
Working memory - *d'*	-	-0.37 (-2.02, 1.29)	0.16 (-0.88, 1.20)	-0.76 (-1.76, 0.23)	3.08 (3) *p*=0.38
Response inhibition - SSRT	-	38.1 (-102.0, 178.3)	16.5 (-54.1, 87.2)	51.0 (-17.6, 119.6)	2.45 (3) *p*=0.48
Emotion recognition - 6 AFC	-	-0.10 (-0.26, 0.07)	-0.03 (-0.13, 0.08)	-0.03 (-0.13, .07)	1.72 (3) *p*=0.64

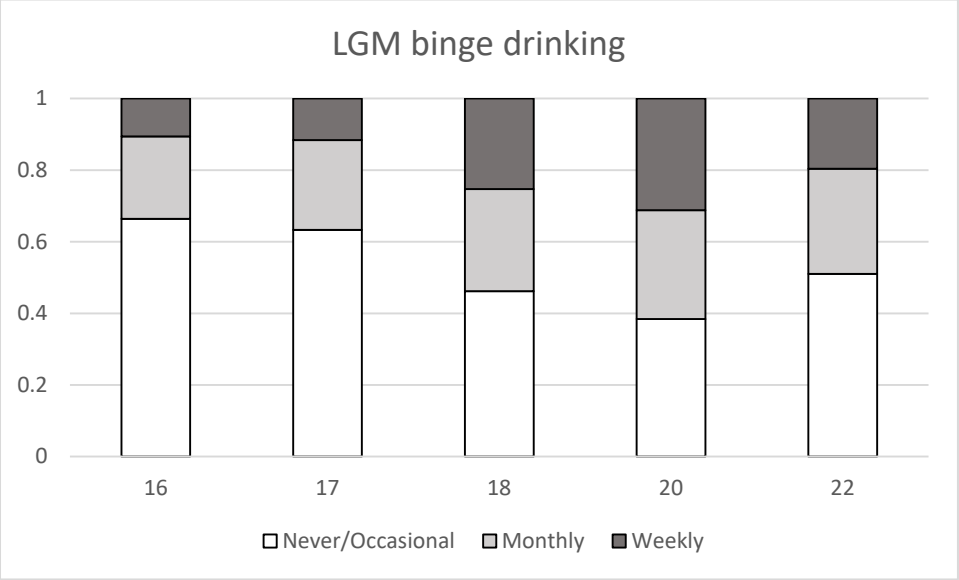

Figure S3. Latent growth model of binge drinking measures from 16 to 22 years (*n*=3,155)

Table S10. Latent growth model (fully adjusted models)

n=3,155	Working memory		Response inhibition		Emotion recognition	
Binge drinking	*b* (95% CI)	*p*	*b* (95% CI)	*p*	*b* (95% CI)	*p*
Intercept	0.03 (-0.03, 0.09)	0.32	0.02 (-0.02, 0.06)	0.25	0.01 (0.00, 0.01)	0.02
Slope	0.53 (0.10, 0.96)	0.02	-0.16 (-0.44, 0.08)	0.26	0.07 (0.02, 0.11)	<0.01

Note: models adjusted for sex, tenure, income, social status, housing tenure, maternal education, maternal smoking in pregnancy, working memory at age ~11 years and head injury/ unconsciousness up to age 11 years, and tobacco and cannabis use up to age 16 years

Table S11. SNPs associated with working memory (*d*' prime) at age 24 in ALSPAC at *p*-value threshold of < $5x10^{-6}$ and clumped for independence, in ascending order of p-value

SNP ID	CHR	Position	Alleles	EAF	b	SE	p
rs35225200	4	103146888	C/A	0.08	-0.27	0.05	7.98×10^{-7}
rs7171755	15	73850580	A/G	0.41	0.14	0.03	2.22×10^{-6}
rs73249722	12	3850230	G/A	0.16	0.17	0.04	4.73×10^{-6}

Note. CHR = Chromosome, EAF = Effect Allele frequency.

Table S12. SNPs associated with emotion recognition ability (total hits) at age 24 in ALSPAC at *p*-value threshold of < $5x10^{-6}$ and clumped for independence, in ascending order of p-value

SNP ID	CHR	Position	Alleles	EAF	b	SE	p
rs72739201	9	85440281	C/T	0.05	-0.32	0.07	1.63×10^{-6}
rs67715018	4	62443938	G/A	0.23	0.16	0.03	2.03×10^{-6}
rs148742906	9	78666445	T/C	0.02	-0.53	0.11	2.13×10^{-6}
rs75207553	5	77445015	T/A	0.04	0.35	0.07	2.15×10^{-6}
rs11253557	10	1043667	T/C	0.20	0.16	0.04	3.89×10^{-6}
rs1958349	14	90834860	T/C	0.73	-0.14	0.03	4.02×10^{-6}

Note. CHR = Chromosome, EAF = Effect Allele frequency.

Table S13. SNPs associated with response inhibition (stop signal reaction time) at age 24 in ALSPAC at *p*-value threshold of < $5x10^{-6}$ and clumped for independence, in ascending order of p-value

SNP ID	CHR	Position	Alleles	EAF	b	SE	p
rs112422339	4	164619499	G/C	0.04	0.37	0.07	4.62×10^{-7}
rs112820797	4	3182738	T/G	0.10	0.24	0.05	1.63×10^{-6}
rs77355006	2	157874790	T/G	0.02	-0.48	0.10	2.13×10^{-6}
rs138035342	8	128146894	A/G	0.03	-0.42	0.09	3.67×10^{-6}
rs12591197	15	35263640	C/A	0.76	0.15	0.03	3.80×10^{-6}
rs147627531	4	20991383	T/C	0.01	-0.61	0.13	3.98×10^{-6}

Note. CHR = Chromosome, EAF = Effect Allele frequency.

Table S14. Tests of the unweighted and weighted regression dilution I^2_{GX}

	I^2_{GX} Unweighted	I^2_{GX} Weighted	mF
Drinks per week > working memory	0.845	0.709	47.515
Drinks per week > response inhibition	0.845	0.711	47.515
Drinks per week > emotion recognition	0.845	0.710	47.515

Note. Unweighted estimates only take into account dilution in the SNP-exposure effects, whereas weighted estimates account for the SE of the SNP-outcome effects (21). mF is the mean F-statistic.

Table S15. Tests of heterogeneity in the SNP-exposure association using the IVW method

	Cochran's Q	**df**	**P-value**
Drinks per week > working memory	121.23	86	0.007
Drinks per week > response inhibition	122.97	86	0.006
Drinks per week > emotion recognition	88.61	86	0.403

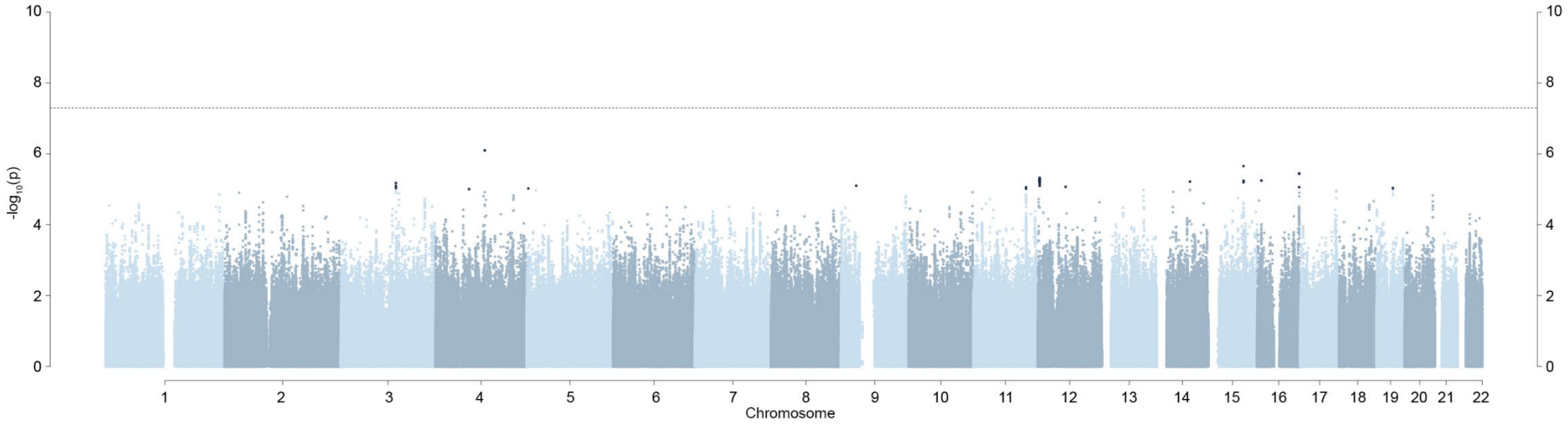

Figure S4. Manhattan plot of a GWAS of working memory (*d*' prime, at age 24) in ALSPAC

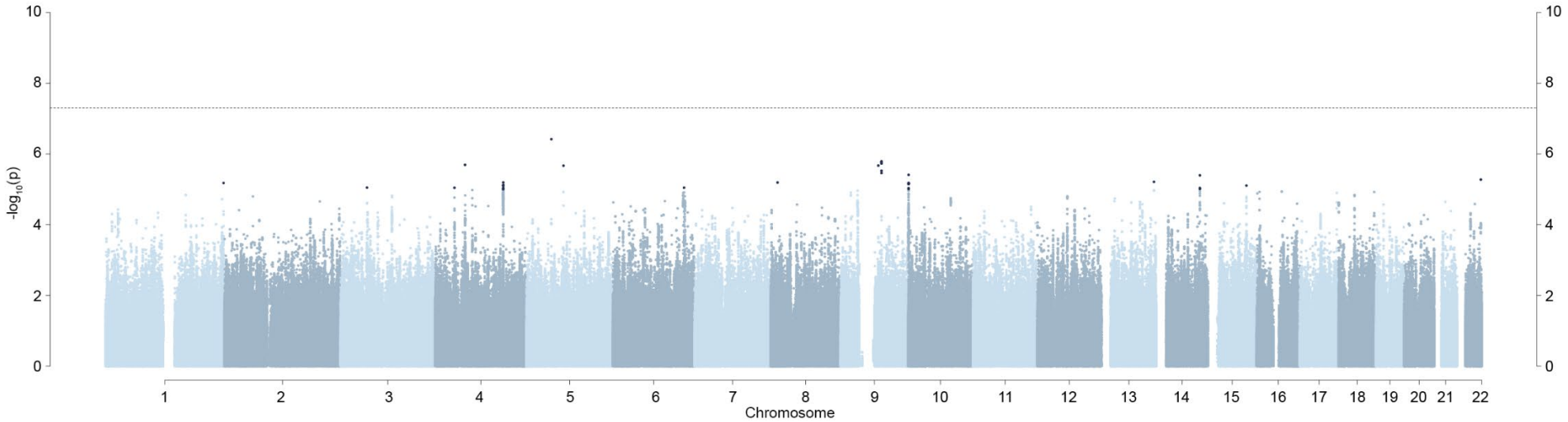

Figure S5. Manhattan plot of a GWAS of emotion recognition ability (total hits, at age 24) in ALSPAC

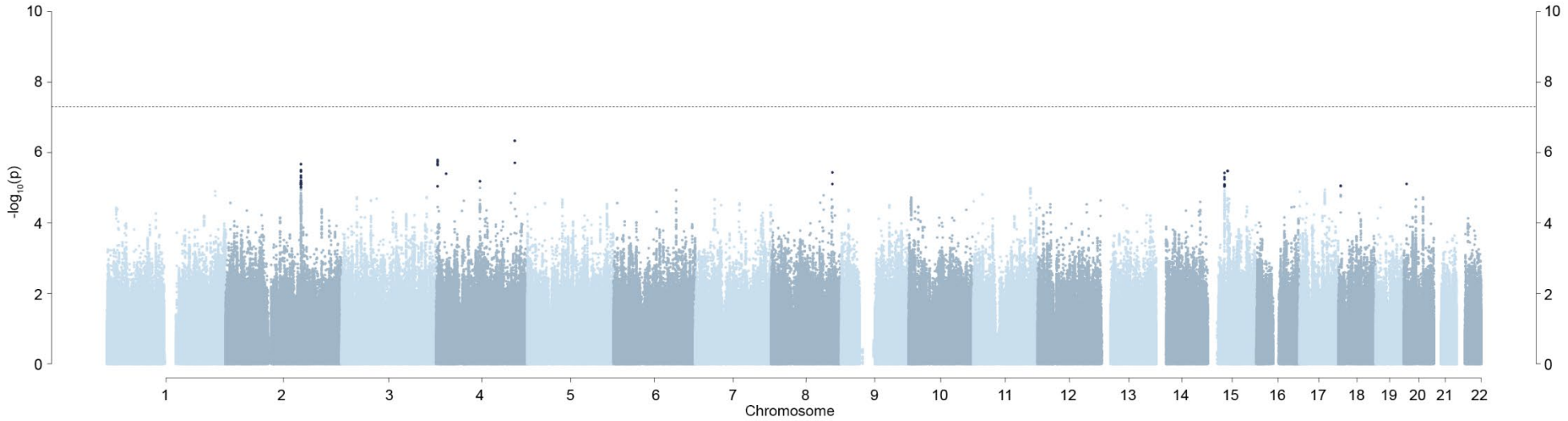

Figure S6. Manhattan plot of a GWAS of response inhibition (stop signal reaction time, at age 24) in ALSPAC

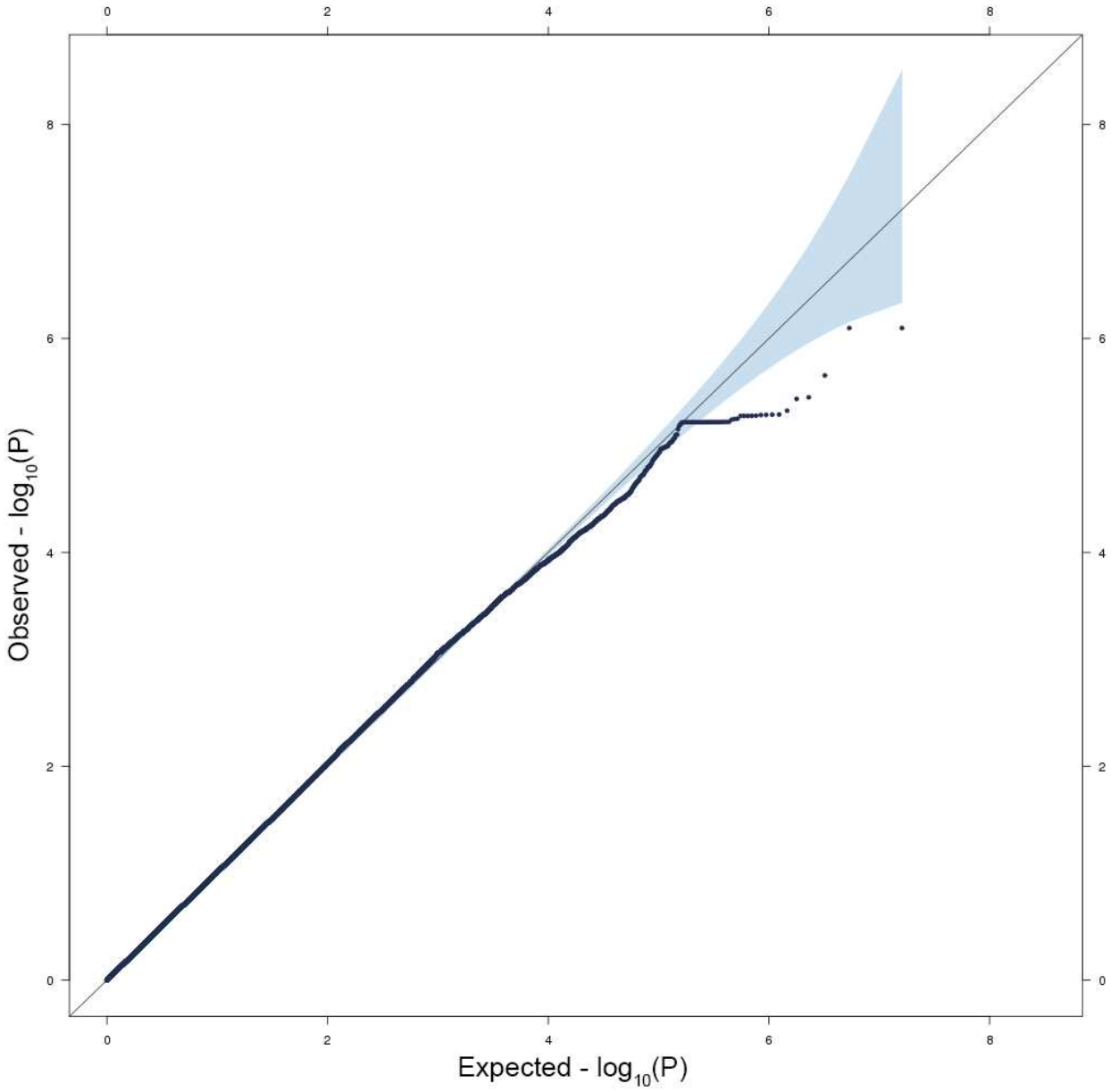

Figure S7. QQ plot of a GWAS of working memory (*d*' prime, at age 24) in ALSPAC (λ = 1.022)

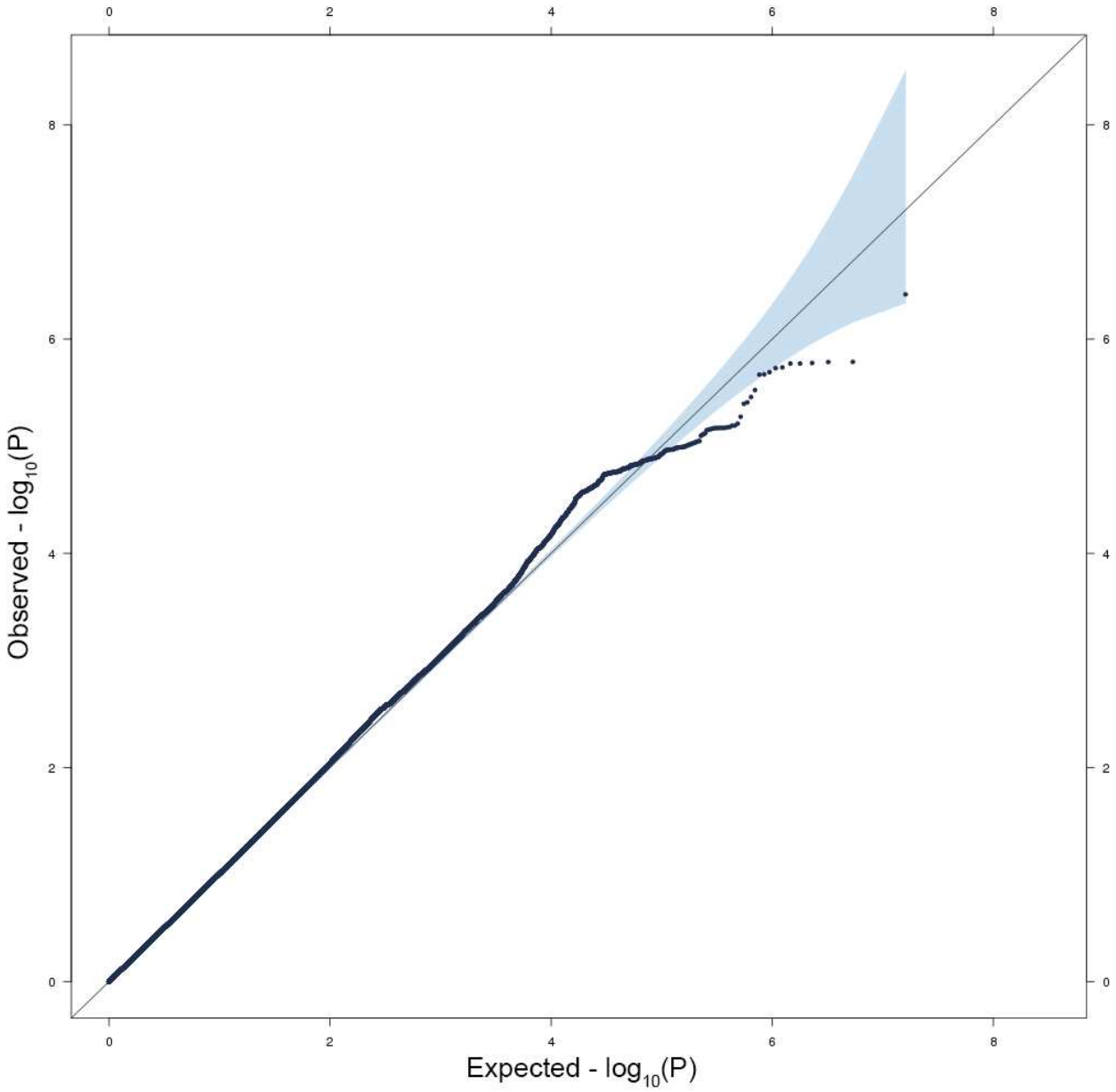

Figure S8. QQ plot of a GWAS of emotion recognition ability (total hits, at age 24) in ALSPAC (λ = 1.004)

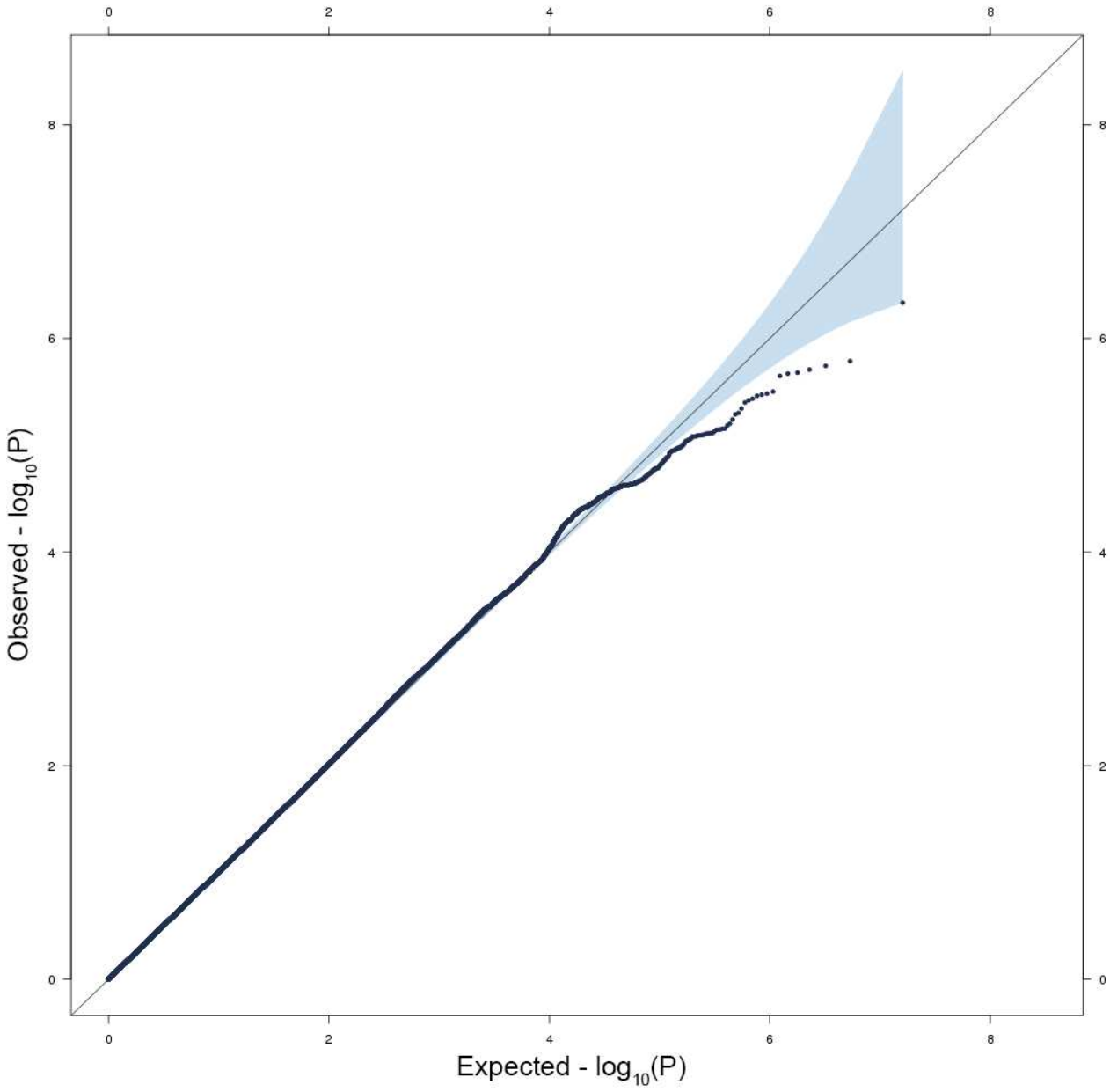

Figure S9. QQ plot of a GWAS of response inhibition (stop signal reaction time, at age 24) in ALSPAC (λ = 1.015)

References

1. McNicol D. A Primer of Signal Detection Theory. Norwich: George Allen & Unwin Ltd; 1972.
2. Stanislaw H, Todorov N. Calculation of signal detection theory measures. Behav Res Methods, Instruments, Comput. 1999;3(I):137–49.
3. Kane MJ, Conway ARA, Miura TK, Colflesh GJH. Working memory, attention control, and the N-back task: a question of construct validity. J Exp Psychol Learn Mem Cogn. 2007;33(3):615–22.
4. Meule A. Reporting and interpreting working memory performance in n-back tasks. Front Psychol. 2017;8:352.
5. Haatveit BC, Sundet K, Hugdahl K, Ueland T, Melle I, Andreassen OA. The validity of d prime as a working memory index: results from the "Bergen n-back task". J Clin Exp Neuropsychol. 2010;32(8):871–80.
6. Schneider W, Eschman A, Zuccolotto A. E-Prime User's Guide. Pittsburgh: Psychology Software Tool Inc.; 2002.
7. Wagner HL. On measuring performance in category judgment studies of nonverbal behavior. J Nonverbal Behav. 1993;17(1):3–28.
8. Case R, Kurland DM, Goldberg J. Operational efficiency and the growth of short-term memory span. J Exp Child Psychol. 1982;33:386–404.
9. Newsome MR, Scheibel RS, Steinberg JL, Troyanskaya M, Sharma RG, Rauch RA, et al. Working memory brain activation following severe traumatic brain injury. Cortex. 2007;43(1):95–111.
10. Jones HJ, Gage SH, Heron J, Hickman M, Lewis G, Munafò MR, et al. Association of combined patterns of tobacco and cannabis use in adolescence with psychotic experiences. JAMA Psychiatry. 2018;75(3):240–6.
11. Sclove L. Application of model-selection criteria to some problems in multivariate analysis. Psychometrika. 1987;52:333–43.
12. Akaike H. Factor analysis and AIC. Psychometrika. 1987;52:317–32.
13. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6:461–4.
14. Lo YB, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. Biometrika. 2001;88(3):767–78.
15. McCutcheon A. Latent Class Analysis. CA: Sage Publications; 1987.
16. Taylor AE, Jones HJ, Sallis H, Euesden J, Stergiakouli E, Davies NM, et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2018;47(4):1207–16.
17. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018 May;7.

18. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25.

19. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40(4):304–14.

20. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985–98.

21. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int J Epidemiol. 2016;45(6):1961–74.

22. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLOS Genet. 2017;13(11):e1007081.