**Supplementary Materials**

*N-back working memory task*

This task has been described previously by Mahedy and colleagues(Mahedy et al., 2019). In this task, participants continuously monitored a series of numbers (0-9) presented on a computer screen and pressed the ‘1’ key if the number was the same as the number presented two trials previously (i.e., 2-back), or the ‘2’ key if it was not. Stimuli numbers were presented in black on a white background with a random spatial jitter of 180 pixels in the y-axis and 200 pixels in the x-axis. Each target was presented for 500 ms, followed by a 3,000 ms window in which to respond. There was an initial practice block which consisted of 12 trials containing two targets. This was followed by an experimental block, consisting of 48 trials, with 8 targets, where the target was the number that was identical to the one presented 2 trials previously. 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′, a signal-detection metric that takes into account both hits and false alarms to derive an overall estimate of signal-detection ability (McNicol, 1972; Stanislaw and Todorov, 1999). d′ was calculated as follows:

d’ = invnorm(hits) – invnorm(false alarms)

*Stop signal task*

This task has been described previously by Mahedy and colleagues(Mahedy et al., 2019). In this task, participants were asked to sit in front of a computer screen with their two index fingers placed on two stimulus keys, one for ‘X’ and one for ‘O’. There were two types of trial, the ‘go’ trials and the stop signal trials. In the ‘go’ 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 key as quickly as possible. On 25% of the trials, a beep is heard (stop signal), randomly after the ‘X’ or ‘O’ appears. Participants were asked not to press the corresponding key when they heard the stop signal/beep 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. There were 32 practice trials and the main task consisted of 256 trials, comprising 4 blocks of 64 trials. Each block of 64 trials consisted of 4 sub-blocks of 16 trials. Each sub-block consisted of 12 trials without a stop-signal and 4 trials with a stop-signal/beep. Mean response times for this were calculated. Four metrics were obtained for the stop signal task: (i) an estimate of stop signal reaction time (SSRT) was calculated and used as a measure of inhibitory control (shorter SSRTs indicate slower inhibition); (ii) ‘go’ reaction time; (iii) ‘go’ accuracy; and (iv) ‘stop’ accuracy. In our study we used the SSRT, which was calculated as follows:

SSRTmed = Go Reaction Timemed – Stop Signal Delaymed

The Stop Signal Delaymed (SSD) was calculated 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%.

*Potential confounders*

We selected potential confounders based on established risk factors for autism spectrum disorder (ASD) and those that may theoretically be predictive of emotion recognition ability. These included child’s sex, measures of socioeconomic status (highest education and social class of mother and partner, income and tenure), mothers smoking and binge drinking, a measure from the Wechsler Intelligence Scale for Children (WISC), head injuries, a measure from the Strengths and Difficulties Questionnaire (SDQ) and the short Mood and Feelings Questionnaire (MFQ), anxiety and earlier ASD symptoms.

Highest social class was based on occupation of the mother and partner from questionnaire data at approximately 32 weeks gestation. Highest education level for mothers and partners was taken from questionnaire data, where participants were asked “What educational qualifications do you, your partner, your mother, and your father have?”. They were asked to select all options that applied to them and we used the highest education qualification for the participant. The options for this were Certificate of Secondary Education (CSE) / none, vocational qualifications, O level (qualifications usually taken at age 16), A level (qualifications usually taken at age 18) or degree. We used the highest from mothers and partners if both were provided. for income (quintiles) and tenure (owned/mortgaged, subsidised renting and private renting) were collected during pregnancy, around the time of enrolment. Measures for mothers smoking (yes/no) and binge drinking (based on self-report of alcohol consumption in the previous week) were obtained when children were aged 12 and 4 via questionnaire.

We used total scores from the WISC (Wechsler et al., 1992), assessed at age 8 in research clinics, as a measure of IQ. A short form of the measure was used, comprising of five verbal and five performance subtests, results of which were combined to give a total score for IQ. We used a total score from the SDQ at age 6 obtained from parent questionnaire data. We used a total score from the short MFQ which participants completed on a computer at a research clinic visit when they were around age 12. The anxiety measure was obtained from child-based questionnaires completed by the mother/caregiver when the child was around age 7, where questions based on DSM-IV criteria were asked and a binary measure of any anxiety disorder was used. Two head injury variables were created, one for earlier head injury (based on questionnaire data up to age 8) and another for later head injury (based on questionnaire data up to age 11). The earlier measure for ASD was also based on questionnaire data obtained at around the age of 7, with the Social and Communication Disorders Checklist (SCDC) as well.

*Multiple Imputation*

To impute the data to use for the emotion recognition task (ERT), SSRT and working memory analyses, we included all relevant outcome, exposure and covariate variables, as well as working memory at age 11, measures of ASD symptoms also using the SCDC at ages 10 and 13, SDQ measures at ages 9, 11 and 13, MFQ measures at ages 10 and 13, the Wechsler Abbreviated Scale of Intelligence (WASI) at age 15, the sociability scale of the Emotionality, Activity, Sociability (EAS) temperament scale at age 3, measures for crowding, parity, mothers cannabis use when their child was aged 9, mothers alcohol and smoking frequency when the child was aged 13.

To impute the data to use for the Diagnostic Analysis of Nonverbal Accuracy (DANVA) analyses, we included all relevant outcome, exposure and covariate variables, as well as measures of ASD symptoms also using the SCDC at ages 10 and 13, SDQ measures at ages 9, 11 and 13, the WASI, the sociability scale of the EAS temperament scale at ages 3, 4 and 5, measures for crowding, parity and mums education.

*Polygenic risk scores*

We constructed thirteen different polygenic risk scores (PRS) based on different thresholds of the phenotypes genome-wide association study (GWAS) p-values (0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.01, 1x10-3, 1x10-4, 1x10-5, 1x10-6, 1x10-7, 5x10-8). For the ERT and DANVA there were no SNPs that passed the more stringent p-value thresholds (1x10-7 for the ERT and 1x10-6 for the DANVA). However, we only present results for those that explained the most variance in the outcome (see supplementary Tables S1-S3 for all results). For ASD this was 1x10-3, for the DANVA this was 0.5 and for the ERT this was also 0.5.

**Supplementary Table S1. Associations between standardised ERT PRS and number of correct responses in the ERT in ALSPAC.**

|  |  |
| --- | --- |
| **P-value threshold for polygenic risk score** | **R2** |
| 0.000001 | 0.008 |
| 0.00001 | 0.09 |
| 0.0001 | 0.45 |
| 0.001 | 0.80 |
| 0.01 | 0.92 |
| 0.05 | 0.94 |
| 0.1 | 0.94 |
| 0.2 | 0.95 |
| 0.3 | 0.95 |
| 0.4 | 0.95 |
| 0.5 | 0.95\* |

*\*Polygenic risk score that explained the most variance in the outcome*

**Supplementary Table S2. Associations between standardised DANVA PRS and number of correct responses in the DANVA in ALSPAC.**

|  |  |
| --- | --- |
| **P-value threshold for polygenic risk score** | **R2** |
| 0.00001 | 0.01 |
| 0.0001 | 0.09 |
| 0.001 | 0.33 |
| 0.01 | 0.55 |
| 0.05 | 0.61 |
| 0.1 | 0.62 |
| 0.2 | 0.63 |
| 0.3 | 0.63 |
| 0.4 | 0.64 |
| 0.5 | 0.64\* |

*\*Polygenic risk score that explained the most variance in the outcome*

**Supplementary Table S3. Associations between standardised ASD polygenic risk score and ASD symptoms in ALSPAC.**

|  |  |
| --- | --- |
| **P-value threshold for polygenic risk score** | **R2** |
| 0.00000005 | 0.00002 |
| 0.0000001 | 0.00002 |
| 0.000001 | 0.0001 |
| 0.00001 | 0.000008 |
| 0.0001 | 0.00006 |
| 0.001 | 0.0002\* |
| 0.01 | 0.0001 |
| 0.05 | 0.0000009 |
| 0.1 | 0.000002 |
| 0.2 | 0.000004 |
| 0.3 | 0.000002 |
| 0.4 | 0.000003 |
| 0.5 | 0.000003 |

*\*Polygenic risk score that explained the most variance in the outcome*

**Supplementary Table S4. Unimputed results for associations between ASD and emotion recognition task outcome for fully adjusted model.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Beta** | **2.5% CI** | **97.5% CI** | **P** |
| -0.19 | -0.31 | -0.08 | 1.15x10-03 |

*Model with number of correct responses in the emotion recognition task as the outcome and ASD symptoms as the exposure adjusted for sex, highest education, highest social class, mothers smoking, mothers’ binge drinking, income, tenure, WISC, head injury, SDQ, MFQ, anxiety, earlier ASD symptoms*

**Supplementary Table S5. Imputed results for associations between ASD and SSRT outcome (N=3,579).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Beta** | **2.5% CI** | **97.5% CI** | **P** |
| Model 1a | 0.27 | -0.35 | 0.89 | 0.39 |
| Model 2b | 0.15 | -0.46 | 0.76 | 0.64 |
| Model 3c | 0.008 | -0.60 | 0.61 | 0.98 |
| Model 4d | 0.02 | -0.59 | 0.62 | 0.96 |
| Model 5e | -0.27 | -0.90 | 0.37 | 0.41 |
| Model 6f | -0.29 | -0.93 | 0.34 | 0.37 |
| Model 7g | -0.22 | -0.92 | 0.46 | 0.51 |

*Models with SSRT as the outcome and ASD symptoms as the exposure were adjusted as follows in addition to covariates in the previous model. a: No covariates, b: sex + highest education + highest social class + mothers smoking + mothers’ binge drinking + income + tenure, c: WISC, d: head injury, e: SDQ, f: MFQ + anxiety, g: earlier ASD symptoms*

**Supplementary Table S6. Imputed results for associations between ASD and working memory outcome (N=3,579).**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Beta** | **2.5% CI** | **97.5% CI** | **P** |
| Model 1a | -0.02 | -0.03 | -0.01 | 3.46x10-07 |
| Model 2b | -0.02 | -0.03 | -0.01 | 1.66x10-06 |
| Model 3c | -0.02 | -0.03 | -0.01 | 2.06x10-05 |
| Model 4d | -0.02 | -0.03 | -0.01 | 1.90x10-05 |
| Model 5e | -0.02 | -0.02 | -0.005 | 0.002 |
| Model 6f | -0.01 | -0.02 | -0.005 | 0.003 |
| Model 7g | -0.02 | -0.03 | -0.005 | 0.004 |

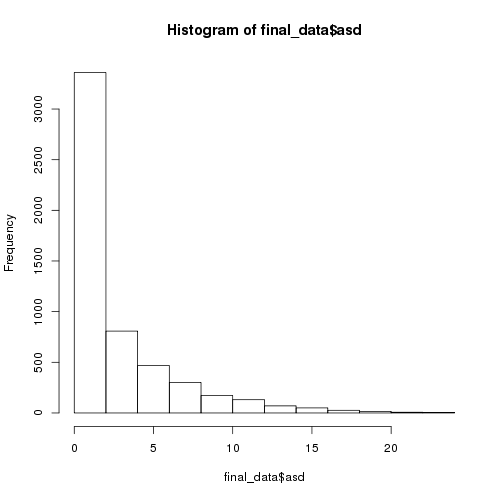
*Models with working memory as the outcome and ASD symptoms as the exposure were adjusted as follows in addition to covariates in the previous model. a: No covariates, b: sex + highest education + highest social class + mothers smoking + mothers’ binge drinking + income + tenure, c: WISC, d: head injury, e: SDQ, f: MFQ + anxiety, g: earlier ASD symptoms*

**Supplementary Table S7. Unimputed results for associations between the DANVA measure and ASD symptoms for fully adjusted model.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Beta** | **2.5% CI** | **97.5% CI** | **P** |
| 0.0003 | -0.05 | 0.05 | 0.99 |

*Models with the ASD symptoms as the outcome and number of correct responses in the DANVA as the exposure adjusted for sex, highest education, highest social class, mothers smoking, mothers’ binge drinking, income, tenure, WISC, head injury, SDQ, anxiety, earlier ASD symptoms*

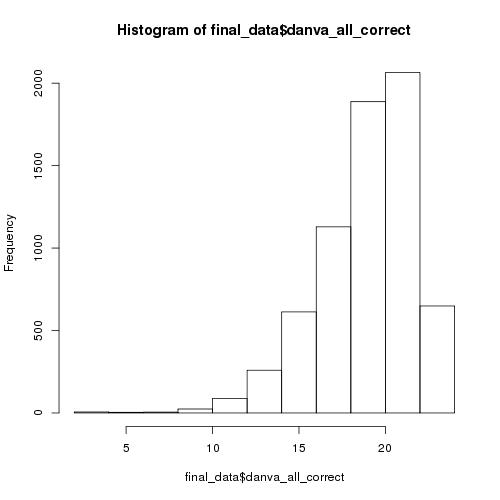
**Supplementary Figure S1. Histogram of SCDC assessed ASD symptoms.**



SCDC score for ASD

*The distribution of SCDC assessed ASD symptoms indicates that these data are zero-skewed.*

**Supplementary Figure S2. Histogram of number of correct responses in the DANVA.**

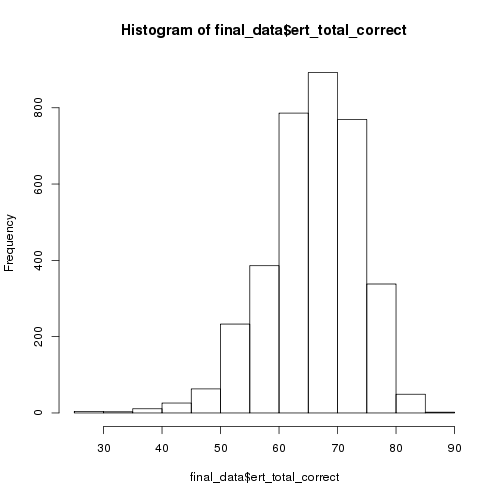


Number of correct responses in the DANVA

*The distribution of the number of correct responses in the DANVA indicates that these data are slightly negatively skewed.*

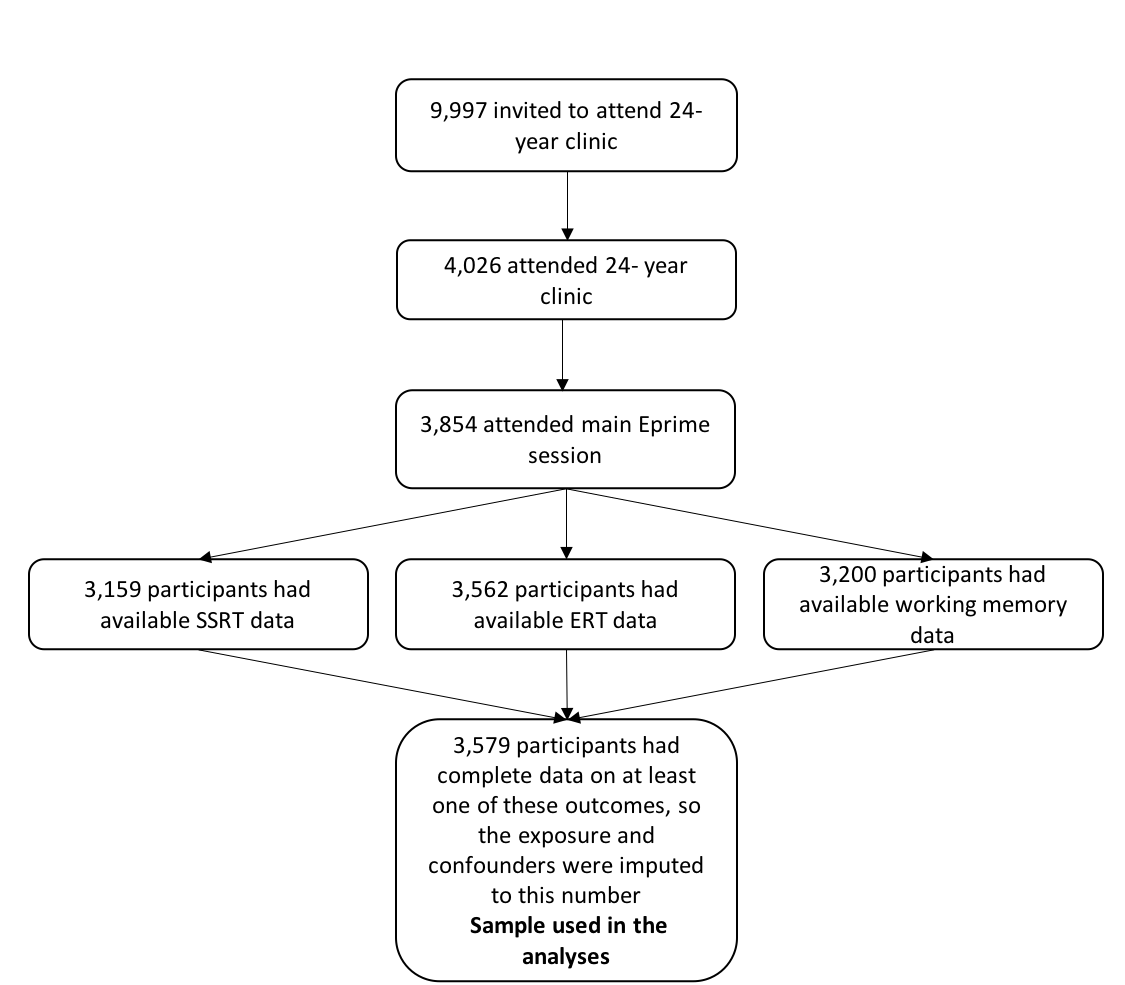
**Supplementary Figure S3. Histogram of number of correct responses in the ERT.**

Number of correct responses in the ERT

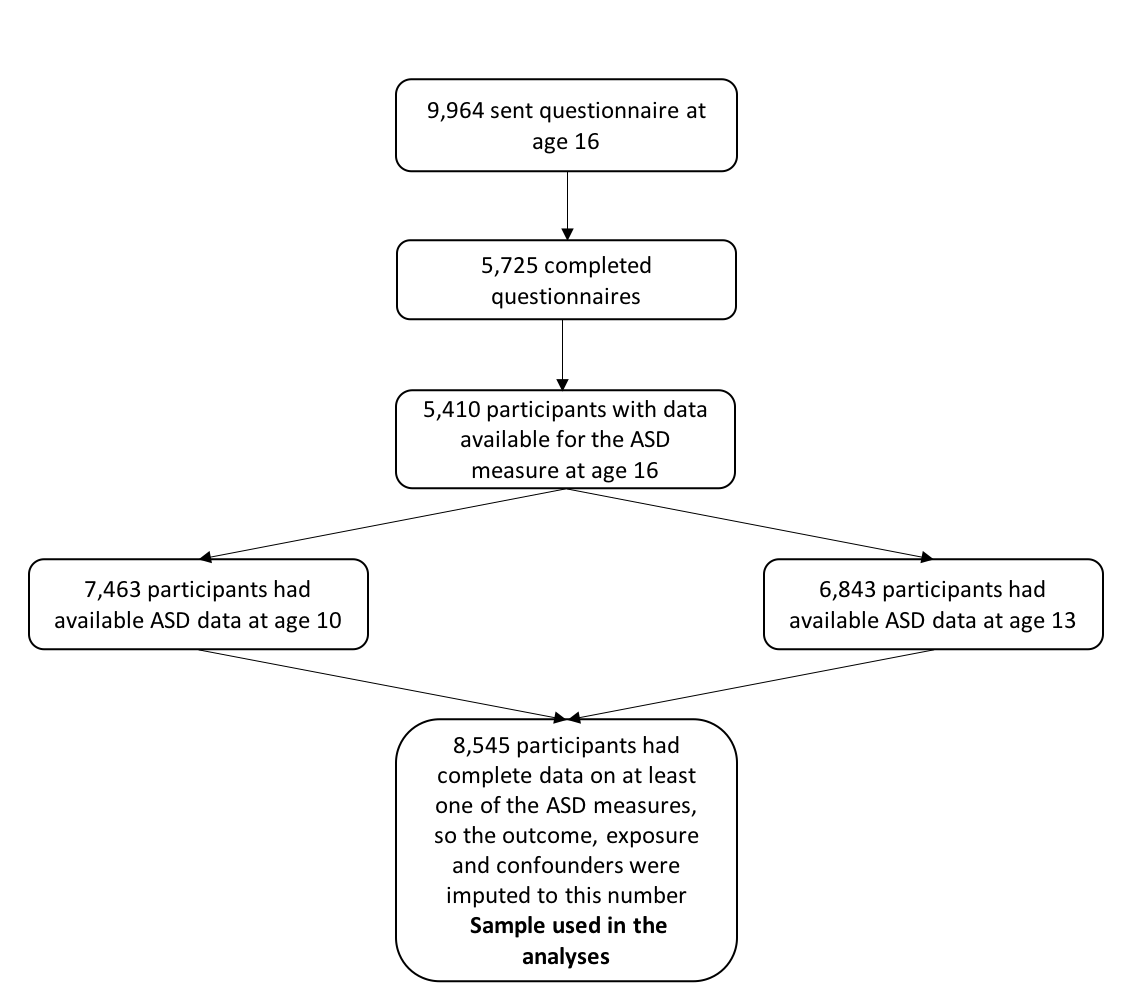


*The distribution of the number of correct responses in the DANVA indicates that these data are slightly negatively skewed.*

**Supplementary Figure S4. Sample attrition in ALSPAC for measures at the age 24 clinic (Study 1).**



**Supplementary Figure S5. Sample attrition in ALSPAC for the ASD outcome at age 16 (Study 2).**



**Supplementary Figure S6. Distribution of imputed variables compared to original data for ERT, SSRT and working memory outcome analyses.**

A close up of a map

Description automatically generated*Density plots for each variable that was imputed for the ERT, SSRT, working memory and ASD analyses. Blue represents the observed data and red represents the imputed data for 100 imputations.*

**Supplementary Figure S7. Distribution of imputed variables compared to original data for ASD outcome analyses.**

A close up of a map

Description automatically generated*Density plots for each variable that was imputed for the DANVA and ASD analyses. Blue represents the observed data and red represents the imputed data for 100 imputations.*

**References**

Mahedy, L., Suddell, S., Skirrow, C., Wootton, R., Fernandes, G.S., Heron, J., Field, M., Hickman, M., Munafo, M.R., 2019. Alcohol use and cognitive functioning in young adults: an observational and Mendelian randomisation study. medRxiv 19003327.

McNicol, D., 1972. A primer of signal detection theory. George Allen & Unwin Ltd, Norwich.

Stanislaw, H., Todorov, N., 1999. Calculation of signal detection theory measures. Behav. Res. Methods, Instruments, Comput. 31, 137–149.

Wechsler, D., Golombok, J., Rust, S., 1992. Manual for the Wechsler Intelligence Scale for Children. Psychological Corporation, Sidcup, UK.