**Supplementary Material**

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

Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study. The initial number of pregnancies enrolled is 14,541 (for these at least one questionnaire has been returned or a "Children in Focus" clinic had been attended by 19/07/99). Of these initial pregnancies, there was a total of 14,676 foetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, the total sample size for data collected after the age of seven is therefore 15,454 pregnancies, resulting in 15,589 foetuses. Of these 14,901 were alive at 1 year of age. Part of this data was collected using REDCap (https://projectredcap.org/resources/citations/). Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool: http://www.bristol.ac.uk/alspac/researchers/our-data/. Further details of the study, measures and sample can be found elsewhere (1-3). Where families included multiple births, we included the oldest sibling.

## Social Communication Disorders Checklist (SCDC)

We investigated possible measurement variance in the SCDC across age and sex in a number of steps. First, we examined associations between the SCDC and additional measures of ASD: results are shown in Supplementary Table 4. Associations by age had largely overlapping confidence intervals with the exceptions of stronger associations for measures closer in time,

specifically (i) stronger associations with childhood ASD diagnosis for the SCDC at age 7 years compared to ages 17 and 25 years, and for the SCDC at age 10 compared to at age 17 years, (ii) stronger associations with high-risk for childhood ASD for the SCDC at ages 7 and 10 years compared to later assessments, and (iii) stronger associations with high-risk for adult ASD according to parent-report for the SCDC at age 25 years compared to earlier assessments. Associations by sex showed overlapping confidence intervals for males and females at all five ages.

We then assessed measurement invariance first by age and then by sex using structural equation modeling to model a latent SCDC factor indexed by the 12 ordinal SCDC items. In-line with recommendations (4), we evaluated increasingly stringent types of measurement invariance: (i) configural invariance (same pattern of free and fixed loadings across age/sex). (ii) metric invariance (similar degree of factor loadings across age/sex), (iii) scalar invariance (similar items thresholds across age/sex), and (iv) residual invariance (similar items residuals across age/sex). Models were fit in Mplus (5) using weighted least square parameter estimates (WLSMV). Model fit was assessed using a variety of indices including the comparative fit index (CFI), root-mean-square error of approximation (RMSEA) and standardized root mean squared residual (SRMSR), for which values of ≥0.95, ≤0.06 and ≤0.08 are generally considered good fit (6). Model fit indices are shown in Supplementary Table 5.

To examine measurement invariance by age we started by fitting a single SCDC factor with factor loadings, SCDC item thresholds and residuals free to vary by age (model A1); this model showed good model fit according to the RMSEA and SRMR although the CFI was <0.95. Fixing factor loadings across ages (model A2) led to an improvement in model fit providing evidence of metric invariance, but subsequently fixing item thresholds by age (model A3) resulted in poorer model fit (CFI=0.93, ΔCFI>0.01) (4, 7) suggesting that the SCDC does not show scalar invariance by age. As scalar invariance was not established, residual invariance was not evaluated. These

models suggest acceptable measurement invariance in that the basic organization of the underlying SCDC construct is supported across these ages (configural invariance), with each SCDC item contributes to a latent SCDC construct to a similar degree across these ages (metric invariance), but that mean differences in the shared variance of these items may not all be captured by mean differences in the latent construct (scalar noninvariance).

We examined measurement invariance across sex using the theta parameterization in Mplus to enable the modelling of residual variances with multiple groups (sex) when using ordinal factor indices (SCDC items). Based on our findings of metric invariance by age, factor loadings were fixed but thresholds and residuals freed across age. We began by fitting the single SCDC factor with factors loadings and thresholds free across sex (model S1) (residual variances were fixed for identification purposes)(5): this model showed good model fit. Fixing factor loadings across sex (model S2) led to an improvement in model fit providing evidence of metric invariance. Fixing item thresholds (enabling the freeing of residual variances) by sex (model S3) retained good model fit and subsequently fixing residual variances (model A4) led to improvement in model fit. These models provide evidence of measurement invariance in the SCDC across males and females.

**Not in Education, Employment or Training (NEET) status**

In-line with the UK Office for National Statistics definition (8) individuals were classified as being in employment if they were in full-time, part-time, irregular/occasional work or self-employed; individuals who were not in employment, doing a modern apprenticeship or other government supported training/work-experience scheme or in full-time education were defined as being NEET and included those doing voluntary work, unable to work through sickness/disability and those who were a full/part-time carer.

**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. Each model was run with 5000 random starting values and 500 optimizations (STARTS = 5000 500 in Mplus) (5). Models were fit for a piecewise growth model with a single intercept and two linear slope factors: one for ages 7, 10, 13 and 17 years and one for ages 17 and 25 years: the second slope variance was fixed to zero to avoid nonidentification as only two time-points were included in this growth factor. Fit statistics are shown in Supplementary Table 6. Model fit significantly improved, as indicated by the fall in loglikelihood value, sample size adjusted Bayesian information criterion, Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest and Bootstrapped Likelihood Ratio Test, from the one- to three-class solutions. The Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest indicated no significant improvement in model fit from the three- to four-class solution: the three-class solution was therefore selected, which showed high classification accuracy (entropy = 0.92).

***Sensitivity analyses: regular parental contact***

Sensitivity analyses limited the sample to those with regular parent-offspring contact at age 25 years, assessed by parent-report as seeing their child at least once a month (N=3326/4482). Fit statistics are shown in Supplementary Table 7. Model fit significantly improved, as indicated by the fall in loglikelihood value, sample size adjusted Bayesian information criterion, Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest and Bootstrapped Likelihood Ratio Test, from the one- to four-class solutions. However, the four-class solution included an additional 'subthreshold' class which did not exceed the SCDC (Social Communication Disorders Checklist) cut-point of ≥9 at any age. The clinical relevance of this additional fourth class was uncertain and thus further analyses checks for the purposes of the current study was restricted to the three-class solution which still showed high classification accuracy (entropy = 0.95). This model of three classes is similar to the model for the full sample (see Supplementary Figure 1), although the declining

class had lower initial levels, which may reflect the impact of missing data by including only those with regular parent-offspring contact (and completed data) at age 25 years. As with the primary model (on all the sample), male sex was associated with an increased likelihood of being in the declining trajectory class (OR=1.83, 95% CI=1.34-2.51, p<0.001), but not the late-emerging class (OR=1.14, 95% CI=0.83-1.56, p=0.42) compared to the low class. Higher parental income was associated with a decreased likelihood of being in the late-emerging (OR=0.92, 95% CI=0.86-0.99, p=0.02) and somewhat the declining (OR=0.94, 95% CI=0.87-1.00, p=0.06) groups compared to the low class, with similar levels of association between the two (declining vs late-emerging OR=1.02, 95% CI=0.93-1.12, p=0.73).

***Sex specific developmental trajectories***

Fit statistics for growth mixture models run separately for males and females are shown in Supplementary Table 8. For males model fit significantly improved, as indicated by the fall in loglikelihood value, sample size adjusted Bayesian information criterion, and Bootstrapped Likelihood Ratio Test, from the one- to four-class solutions, however the Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest indicated no significant improvement in model fit from the three- to four-class solution: the three-class solution was therefore selected, which showed high classification accuracy (entropy = 0.94). For females, model fit significantly improved, as indicated by the fall in loglikelihood value, sample size adjusted Bayesian information criterion, and Bootstrapped Likelihood Ratio Test, from the one- to three-class solutions, however the Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest indicated no significant improvement in model fit from the two- to three-class solution: the two-class solution was therefore selected, which showed high classification accuracy (entropy = 0.92) – this model did not include a declining class. Sex-specific models are shown in Supplementary Figure 2. For males, higher parental income was associated with a decreased likelihood of being in the late-emerging (OR=0.90, 95% CI=0.84-0.97, p=0.003) and the declining (OR=0.89, 95% CI=0.83-0.95, p=0.001) groups compared to the low class, with similar levels of association between the two (declining vs late-

emerging: OR=0.99, 95% CI=0.90-1.09, p=0.74). For females, higher parental income was associated with a decreased likelihood of being in the late-emerging (OR=0.89, 95% CI=0.85-0.95, p<0.001) group compared to the low class.

***Deriving developmental trajectories with varying levels of missingness***

Sensitivity analyses were also conducted deriving trajectories with varying levels of missingness. While primary analyses required at least 2 time-points of SCDC data, we re-ran sensitivity analyses requiring at least 1, 3, 4 and 5 time-points. All models used full information maximum likelihood estimation (FIML) which assumes that data are missing at random (or missing completely at random) conditional on the variable in the model: models with more stringent inclusion criteria are likely to be at increased risk of bias, arising from increasing differences between missing and non-missing values. Fit statistics are shown in Supplementary Table 9, which generally showed a similar pattern to those observed in the primary analyses with the exception that for some levels of missingness, a four-class rather than three-class solution may be optimal. However, unlike the composition of the three classes for the three-class solution (the one selected for the main analyses), which was fairly consistent across different level of missingness (see Supplementary Figure 3), the composition of the fourth class of the four-class solution varied. When using more lenient inclusion criteria (requiring at least 1 or 2 SCDC time-points) the 'fourth' class captured those with high-persistent symptoms. When a more stringent inclusion criterion was used (requiring more than 2 SCDC time-points), the additional fourth class was composed of 'intermediate' symptom levels: see Supplementary Figure 3. This likely reflects non-random attrition, whereby individuals with high, persistent ASD symptoms are more likely to drop-out of the study (9). In summary, the three-class solution composition (used in the main analyses) was consistent across varying levels of missingness, while the 'fourth' showed a different composition depending on levels of missingness.

**Missing covariate data**

The primary sample included individuals with at least two time-points of SCDC data (N=8094). Associations between ASD diagnosis in childhood and the availability of our primary measure across development (i.e. of Social Communication Disorders Checklist data not being missing) are shown in Supplementary Table 10. ASD diagnosis in childhood did not show strong association with inclusion in our primary sample, or with missing data in childhood or adulthood, although there was some evidence of an association with an increased likelihood of having missing adolescent data.

The primary investigation into associations between social communication trajectories and associated features (other measures of ASD, IQ and communication problems, peer problems and adult functioning), or 'covariates' were conducted where data were available (N=3376-8057). Sensitivity analyses to examine potential bias arising from missing data were conducted using a range of alternative approaches in Stata[34] using the 'best guess' trajectory classes. Using the 'best guess' trajectory does not account for measurement error in class assignment, however entropy values approaching one (here, model entropy = 0.93) indicate clear allocation of classes and therefore low measurement error in class assignment. Missing data sensitivity analyses included:

i) *Including those with complete cases (CC)*. Including individuals with complete data on all associated features ('covariates') (N=1582).

ii) *Using inverse probability weighting (IPW)*.(10) Weights were derived from a logistic regression analysis of covariate data for a set of measures assessed in or soon after pregnancy with minimal missingness that were that were associated with the presence of complete-case data (1582/8094) (shown in Supplementary Table 11). Missing data on indicators used to derive weights were singly imputed as the modal or mean value from the "full" APSLAC sample. The Hosmer-Lemeshow test was used assess the fit of the missingness model; results did not indicate poor fit (Hosmer-Lemeshow $\chi 2(8)=10.83$, $p=0.21$). IPW was

used in analysis for those with complete data on all cariavtes (N=1582); weights ranged from 1.83 to 34.10.

iii) *Using multiple imputation (MI)* (11). MI by chained equations including those in the primary sample (N=8094). The model included variables used in the IPW analysis, variables used to specify the trajecories and variables included in the covariate analyses (shown in Supplementary Tables 11-13 respectively) as well as 'best guess' trajectory class. The model was used to generate 250 imputed datasets – this was estimated to be sufficient to ensure relatively stable standard errors if the data were imputed again (the recommended 2-stage quadratic rule based on the initial imputation of 250 datasets suggested 109 imputations were needed) (12). Estimates were combined across imputed datasets using Rubin's rules (11).

iv) *Using MI combined with IPW (IPW/MI)* (13). MI by chained equations were also imputed for the primary sample (as above), using IPW to weight the sample to the "full" ALSPAC sample (i.e. including those without at least two time-points of SCDC data) (N=14692). Weights were derived using the same procudere as the IPW-only analyses above, using measures assessed in or soon after pregnancy with minimal missingness that were that were associated with the presence inclusion in the primary sample (8094/14692) (shown in Supplementary Table 14). The Hosmer-Lemeshow test was used assess the fit of the missingness model; results did not indicate poor fit (Hosmer-Lemeshow $\chi2(8)=12.45$, $p=0.132$). Weights ranged from 1.09 to 7.75. The model was used to generate 250 imputed datasets (105 were recommended based on the initial imputation of 250 datasets (12)).

Analyses using those four alternative approaches, as well as the original estimates, as shown in (Supplementary Figures 4-8). Analyses revealed a similar pattern of results across the different approaches, although with much wider confidence intervals for CC and IPW analyses than the orginal analyses. One exception was that the CC and IPW analyses suggested similar levels of childhood peer problems for the late-emerging and declining classes (consistent with impairment being present in childhood for the late-emerging group).

**Supplementary Table 1**. Correlations between Social Communication Disorders Checklist at different ages

| | Age 7 years | Age 10 years | Age 13 years | Age 17 years | Age 25 years |
|---|---|---|---|---|---|
| Age 7 years | 1 | | | | |
| Age 10 years | 0.67 | 1 | | | |
| Age 13 years | 0.54 | 0.63 | 1 | | |
| Age 17 years | 0.42 | 0.49 | 0.60 | 1 | |
| Age 25 years | 0.37 | 0.43 | 0.50 | 0.48 | 1 |

Sample including those with at least 2 time-points of SCDC data: maximum N=8094

**Table 2**. Individual Social Communication Disorders Checklist (SCDC) item frequencies by age

| | Age 7 years | Age 10 years | Age 13 years | Age 17 years | Age 25 years |
|---|---|---|---|---|---|
| 1. Not aware of other people's feelings | 16.81 | 16.24 | 20.77 | 26.32 | 14.08 |
| | (2.26) | (1.72) | (1.90) | (2.48) | (2.47) |
| 2. Does not realise when others are upset or angry | 11.89 | 11.89 | 14.76 | 19.06 | 9.39 |
| | (2.66) | (2.17) | (2.65) | (3.03) | (1.63) |
| 3. Does not notice the effect of his/her behaviour on | 27.58 | 26.50 | 30.83 | 37.92 | 18.96 |
| other members of the family | (2.99) | (2.92 | (3.90) | (6.00) | (3.31) |
| 4. Behaviour often disrupts family life | 19.52 | 17.86 | 21.03 | 24.08 | 13.28 |
| | (2.50) | (2.37) | (2.80) | (5.21) | (2.42) |
| 5. Very demanding of other people's time | 26.25 | 18.35 | 15.95 | 16.98 | 10.9 |
| | (3.59) | (2.87) | (2.03) | (2.96) | (2.10) |
| 6. Difficult to reason with when upset | 37.50 | 34.99 | 36.6 | 38.49 | 24.69 |
| | (5.00) | (4.89) | (5.40) | (7.83) | (6.05) |
| 7. Does not seem to understand social skills e.g. | 19.59 | 13.78 | 11.58 | 10.82 | 5.48 |
| persistently interrupts conversations | (2.53) | (2.14) | (2.06) | (1.75) | (1.19) |
| 8. Does not pick up on body language | 17.65 | 16.42 | 17.45 | 17.22 | 10.15 |
| | (2.20) | (1.96) | (2.02) | (2.32) | (1.66) |
| 9. Does not appear to understand how to behave when | 10.49 | 5.93 | 4.42 | 5.74 | 2.73 |
| out (e.g. in shops, other people's homes) | (1.59) | (1.02) | (0.80) | (3.26) | (0.79) |
| 10. Does not realise if s/he offends people with her/his | 14.84 | 12.30 | 13.83 | 12.11 | 9.82 |
| behaviour | (1.98) | (1.43) | (1.40) | (1.27) | (1.45) |
| 11. Does not respond when told to do something | 36.97 | 28.57 | 29.98) | 26.81 | 9.04 |
| | (2.95) | (2.02) | (2.61 | (2.48) | (1.15) |
| 12. Cannot follow a command unless it is carefully | 7.44 | 7.02 | 7.16 | 7.51 | 5.47 |
| worded | (1.36) | (1.43) | (1.31 | (1.18) | (1.31) |

*Item endorsed quite/sometimes or very/often true (very/often true only in parentheses).

Sample including those with at least 2 time-points of SCDC data: maximum N=8094.

**Supplementary Table 3.** Comparison of ASD and communication subscales by trajectory class

| | Low | | Declining | | Late-emerging | | Declining vs low class | | Late-emerging vs declining | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | (SE) | Mean | (SE) | Mean | (SE) | $\chi^2_{(df=1)}$ | p | $\chi^2_{(df=1)}$ | p |
| Task-based indicator of ASD: age 13 years | | | | | | | | | | |
| Theory of mind | 57.36 | (0.11) | 55.72 | (0.56) | 56.88 | (0.58) | 8.23 | 0.004 | 1.93 | 0.17 |
| Parent-rated ASD "social-behavior" traits: age 25 years | | | | | | | | | | |
| Social skills | 16.98 | (0.08) | 17.24 | (0.45) | 20.77 | (0.41) | 38.44 | <0.001 | 30.67 | <0.001 |
| Routine | 9.87 | (0.30) | 9.59 | (0.26) | 11.86 | (0.24) | 68.23 | <0.001 | 38.31 | <0.001 |
| Switching | 8.54 | (0.04) | 9.20 | (0.23) | 11.57 | (0.25) | 121.80 | <0.001 | 45.33 | <0.001 |
| Imagination | 16.29 | (0.07) | 17.51 | (0.41) | 20.05 | (0.41) | 76.56 | <0.001 | 18.05 | <0.001 |
| Parent-rated ASD "attention to detail" traits: age 25 years | | | | | | | | | | |
| Numbers/patterns | 9.71 | (0.06) | 10.05 | (0.30) | 10.64 | (0.32) | 33.65 | <0.001 | 1.65 | 0.20 |
| Self-rated ASD "social-behavior" traits: age 25 years | | | | | | | | | | |
| Social skills | 14.41 | (0.06) | 19.06 | (0.52) | 18.91 | (0.45) | 15.47 | <0.001 | 18.91 | 0.45 |
| Routine | 7.45 | (0.04) | 9.87 | (0.30) | 10.12 | (0.23) | 3.91 | 0.05 | 0.42 | 0.52 |
| Switching | 6.61 | (0.04) | 9.73 | (0.29) | 9.51 | (0.26) | 16.94 | <0.001 | 0.30 | 0.58 |
| Imagination | 13.88 | (0.06) | 18.05 | (0.47) | 17.70 | (0.45) | 13.50 | <0.001 | 0.26 | 0.61 |
| Self-rated ASD "attention to detail" traits: age 25 years | | | | | | | | | | |
| Numbers/patterns | 8.27 | (0.05) | 10.29 | (0.36) | 10.87 | (0.34) | 2.52 | 0.11 | 1.29 | 0.56 |
| Parent-rated communication problems: age 25 years | | | | | | | | | | |
| Language structure | 1.10 | (0.04) | 3.55 | (0.35) | 8.06 | (0.75) | 48.13 | <0.001 | 27.91 | <0.001 |
| Pragmatic skills | 0.51 | (0.03) | 3.88 | (0.52) | 10.28 | (0.89) | 42.19 | <0.001 | 36.31 | <0.001 |
| Social engagement | 4.91 | (0.15) | 15.58 | (1.33) | 31.61 | (1.53) | 63.07 | <0.001 | 58.14 | <0.001 |

ASD subscales originally informed by factor analyses (14). Theory of mind late-onset versus low $\chi^2(1)=0.66$, $p=0.42$; late-emerging class higher than the low class on all age 25 subscales at $p<0.01$.

**Supplementary Table 4.** Associations between the Social Communication Disorders Checklist (SCDC) and other measures of ASD

| | 7 years | | 10 years | | 13 years | | 17 years | | 25 years | |
|---|---|---|---|---|---|---|---|---|---|---|
| | OR | (95% CI) | OR | (95% CI) | OR | (95% CI) | OR | (95% CI) | OR | (95% CI) |
| Whole sample | | | | | | | | | | |
| Childhood ASD diagnosis | 1.38 | (1.32, 1.44) | 1.36 | (1.30, 1.42) | 1.30 | (1.24, 1.36) | 1.17 | (1.10, 1.25) | 1.24 | (1.18, 1.30) |
| High-risk for childhood ASD | 1.27 | (1.24, 1.29) | 1.23 | (1.21, 1.27) | 1.18 | (1.16, 1.20) | 1.14 | (1.12, 1.17) | 1.17 | (1.15, 1.20) |
| High-risk for adult ASD: parent-rated | 1.20 | (1.17, 1.24) | 1.23 | (1.20, 1.26) | 1.23 | (1.20, 1.26) | 1.18 | (1.15, 1.22) | 1.35 | (1.32, 1.39) |
| High-risk for adult ASD: self-rated | 1.08 | (1.05, 1.11) | 1.09 | (1.06, 1.12) | 1.09 | (1.06, 1.12) | 1.07 | (1.04, 1.10) | 1.12 | (1.09, 1.16) |
| Males | | | | | | | | | | |
| Childhood ASD diagnosis | 1.35 | (1.28, 1.42) | 1.32 | (1.26, 1.39) | 1.29 | (1.22, 1.36) | 1.15 | (1.07, 1.24) | 1.23 | (1.16, 1.30) |
| High-risk for childhood ASD | 1.25 | (1.22, 1.27) | 1.22 | (1.19, 1.25) | 1.18 | (1.15, 1.20) | 1.14 | (1.11, 1.17) | 1.18 | (1.14, 1.21) |
| High-risk for adult ASD: parent-rated | 1.19 | (1.15, 1.23) | 1.22 | (1.18, 1.26) | 1.22 | (1.18, 1.27) | 1.19 | (1.15, 1.23) | 1.38 | (1.32, 1.43) |
| High-risk for adult ASD: self-rated | 1.05 | (1.01, 1.09) | 1.07 | (1.03, 1.11) | 1.07 | (1.03, 1.11) | 1.04 | (0.99, 1.08) | 1.07 | (1.02, 1.13) |
| Females | | | | | | | | | | |
| Childhood ASD diagnosis | 1.40 | (1.24, 1.57) | 1.43 | (1.27, 1.60) | 1.29 | (1.16, 1.43) | 1.24 | (1.10, 1.40) | 1.25 | (1.10, 1.41) |
| High-risk for childhood ASD | 1.29 | (1.24, 1.33) | 1.24 | (1.19, 1.28) | 1.19 | (1.15, 1.23) | 1.16 | (1.13, 1.20) | 1.18 | (1.14, 1.23) |
| High-risk for adult ASD: parent-rated | 1.21 | (1.15, 1.27) | 1.22 | (1.16, 1.28) | 1.22 | (1.17, 1.28) | 1.17 | (1.12, 1.23) | 1.34 | (1.28, 1.39) |
| High-risk for adult ASD: self-rated | 1.10 | (1.06, 1.15) | 1.10 | (1.06, 1.14) | 1.11 | (1.07, 1.15) | 1.10 | (1.07, 1.14) | 1.16 | (1.12, 1.21) |

SCDC as the exposure and other measures of ASD regardless of age for comparability.

**Supplementary Table 5.** Tests of measurement invariance across age and sex

| Model | Free parameters | CFI | RMSEA (90% CI) | SRMR | vs. | Δ parameters | ΔCFI | ΔRMSEA | ΔSRMR | Decision |
|---|---|---|---|---|---|---|---|---|---|---|
| Assessing measurement invariance by age | | | | | | | | | | |
| A1: Configural invariance | 190 | 0.93 | 0.03 (0.03-0.03) | 0.06 | - | | | - | - | - |
| A2: Metric invariance | 146 | 0.94 | 0.03 (0.03-0.03) | 0.07 | A1 | 44 | 0.007 | -0.002 | 0.008 | Accept |
| A3: Scalar invariance | 50 | 0.93 | 0.03(0.03-0.03) | 0.07 | A2 | 96 | -0.014 | 0.003 | 0.011 | Reject |
| Assessing measurement invariance by sex | | | | | | | | | | |
| S1: Configural invariance | 292 | 0.95 | 0.03 (0.03-0.03) | 0.07 | | | | | | |
| S2: Metric invariance | 281 | 0.95 | 0.03 (0.03-0.03) | 0.07 | S1 | 11 | 0.005 | -0.001 | 0.000 | Accept |
| S3: Scalar invariance | 226 | 0.95 | 0.03(0.03-0.03) | 0.07 | S2 | 55 | -0.007 | 0.002 | -0.003 | Accept |
| S4: Residual invariance | 166 | 0.95 | 0.03 (0.03-0.03) | 0.07 | S3 | 60 | 0.005 | -0.002 | 0.003 | Accept |

**Supplementary Table 6**. Model fit indices for growth mixture models

| | LL | Free parameters | ssaBIC | Smallest class | Entropy | VLMR-LRT p value | BLRT p value |
|---|---|---|---|---|---|---|---|
| 1 class | -78146.33 | 11 | 156356.69 | | | | |
| 2 classes | -76409.75 | 15 | 152906.81 | 8.05% (N=651) | 0.94 | <0.001 | <0.001 |
| 3 classes* | -75332.41 | 19 | 150775.41 | 4.99% (N=403) | 0.93 | 0.005 | <0.001 |
| 4 classes | -74685.97 | 23 | 149505.83 | 1.76% (N=142) | 0.92 | 0.053 | <0.001 |

LL=Loglikelihood; ssa= sample size adjusted; BIC= Bayesian Information Criteria; VLMR-LRT=Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest; BLRT=Bootstrapped Likelihood Ratio Test. *Final model.

**Supplementary Table 7**. Sensitivity analyses: model fit indices for growth mixture models for those with regular parent contact

| | LL | Free parameters | ssaBIC | Smallest class | Entropy | VLMR-LRT p value | BLRT p value |
|---|---|---|---|---|---|---|---|
| 1 class | -36990.23 | 11 | 74034.71 | | | | |
| 2 classes | -36124.93 | 15 | 72323.83 | 7.41% (N=246) | 0.97 | <0.001 | <0.001 |
| 3 classes* | -35445.08 | 19 | 70983.87 | 5.90% (N=196) | 0.95 | 0.001 | <0.001 |
| 4 classes | -35036.13 | 23 | 70185.69 | 3.54% (N=118) | 0.95 | 0.011 | <0.001 |

LL=Loglikelihood; ssa= sample size adjusted; BIC= Bayesian Information Criteria; VLMR-LRT=Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest; BLRT=Bootstrapped Likelihood Ratio Test. *Final model.

**Supplementary Table 8**. Model fit indices for growth mixture models by sex

| | LL | Free parameters | ssaBIC | Smallest class | Entropy | VLMR-LRT p value | BLRT p value |
|---|---|---|---|---|---|---|---|
| *Males* | | | | | | | |
| 1 class | -39538.14 | 11 | 79132.74 | | | | |
| 2 classes | -38631.13 | 15 | 77339.25 | 8.64% (N=351) | 0.95 | 0.031 | <0.0001 |
| 3 classes* | -38117.99 | 19 | 76333.50 | 5.72% (N=233) | 0.94 | 0.006 | <0.0001 |
| 4 classes | -37768.15 | 23 | 75654.35 | 2.53% (N=103) | 0.92 | 0.436 | <0.0001 |
| *Females* | | | | | | | |
| 1 class | -38301.35 | 11 | 76659.05 | | | | |
| 2 classes* | -37454.98 | 15 | 74986.80 | 8.07% (N=325) | 0.93 | 0.005 | <0.001 |
| 3 classes | -36956.50 | 19 | 74010.35 | 6.12% (N=247) | 0.82 | 0.071 | <0.001 |

LL=Loglikelihood; ssa= sample size adjusted; BIC= Bayesian Information Criteria; VLMR-LRT=Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest; BLRT=Bootstrapped Likelihood Ratio Test. *Final model.

**Supplementary Table 9**. Model fit indices for growth mixture models with varying levels of missingness

| | LL | Free parameters | ssaBIC | Smallest class | Entropy | VLMR-LRT p value | BLRT p value |
|---|---|---|---|---|---|---|---|
| *1+ data-points: N=9715* | | | | | | | |
| 1 class | -82898.02 | 11 | 165862.09 | | | | |
| 2 classes | -80919.56 | 15 | 161929.17 | 7.43% (N=721) | 0.95 | <0.001 | <0.001 |
| 3 classes | -79727.66 | 19 | 159569.38 | 5.26% (N=511) | 0.91 | 0.005 | <0.001 |
| 4 classes | -78977.00 | 23 | 158092.08 | 2.07% (N=201) | 0.90 | 0.028 | <0.001 |
| *2+ data-points: N=8094 (primary analyses)* | | | | | | | |
| 1 class | -78146.33 | 11 | 156356.69 | | | | |
| 2 classes | -76409.75 | 15 | 152906.81 | 8.05% (N=651) | 0.94 | <0.001 | <0.001 |
| 3 classes* | -75332.41 | 19 | 150775.41 | 4.99% (N=403) | 0.93 | 0.005 | <0.001 |
| 4 classes | -74685.97 | 23 | 149505.83 | 1.76% (N=142) | 0.92 | 0.053 | <0.001 |
| *3+ data-points: N=6614* | | | | | | | |
| 1 class | -69901.46 | 11 | 139864.74 | | | | |
| 2 classes | -68425.81 | 15 | 136935.90 | 8.13% (N=537) | 0.95 | <0.001 | <0.001 |
| 3 classes* | -67494.80 | 19 | 135096.36 | 5.58% (N=369) | 0.94 | 0.079 | <0.001 |
| 4 classes | -66885.55 | 23 | 133900.35 | 3.74% (N=248) | 0.91 | 0.412 | <0.001 |
| *4+ data-points: N=5127* | | | | | | | |
| 1 class | -58073.50 | 11 | 116206.01 | | | | |
| 2 classes | -56854.41 | 15 | 113789.29 | 7.32% (N=375) | 0.96 | <0.001 | <0.001 |
| 3 classes* | -56011.02 | 19 | 112123.97 | 5.51% (N=283) | 0.94 | 0.002 | <0.001 |
| 4 classes | -55433.92 | 23 | 110991.23 | 3.26% (N=167) | 0.92 | 0.027 | <0.001 |
| *5 data-points: N=3021 (complete cases)* | | | | | | | |
| 1 class | -36518.20 | 11 | 116206.01 | | | | |
| 2 classes | -35727.20 | 15 | 113789.29 | 6.00% (N=181) | 0.98 | 0.007 | <0.001 |
| 3 classes* | -35110.48 | 19 | 112123.97 | 4.40% (N=133) | 0.95 | <0.001 | <0.001 |
| 4 classes | -34639.53 | 23 | 110991.23 | 3.03% (N=91) | 0.96 | 0.204 | <0.001 |

LL=Loglikelihood; ssa= sample size adjusted; BIC= Bayesian Information Criteria; VLMR-LRT=Vuong-Lo-Mendell-Rubin Likelihood Ratio Rest; BLRT=Bootstrapped Likelihood Ratio Test. *Final model.

**Supplementary Table 10.** Associations between ASD diagnosis in childhood and missing Social Communication Disorders Checklist (SCDC) data

| | OR | (95% CI) | p |
|---|---|---|---|
| SCDC data available: age 7 years | 1.51 | (0.95-2.39) | 0.08 |
| SCDC data available: age 10 years | 1.00 | (0.65-1.54) | 0.99 |
| SCDC data available: age 13 years | 0.71 | (0.46-1.10) | 0.13 |
| SCDC data available: age 17 years | 0.67 | (0.42-1.07) | 0.09 |
| SCDC data available: age 25 years | 0.92 | (0.57-1.48) | 0.74 |
| Number of SCDC time-points available | 1.25 | (0.77-2.03) | 0.36 |
| Inclusion in primary sample (>1 SCDC time-point) | 0.96 | (0.62-1.48) | 0.85 |

N=13,768 (those with ASD diagnosis data): ASD diagnosis in childhood prevalence = 0.60% (83/13768)

**Supplementary Table 11.** Associations between variables include in the inverse probability weights and missing covariate data

| | Exposure proportion (%) or mean (SE) | | Association with missingness | |
|---|---|---|---|---|
| | Complete covariate data | Incomplete covariate data | Univariable association | Multivariable associations from IPW model* |
| Original enrolment** | 100% | 95.27% (0.26) | - | - |
| Male sex | 35.90% (1.20) | 53.72% (0.62) | OR=2.07, 95% CI=1.84-2.32 | OR=2.17, 95% CI=1.93-2.44 |
| Social disadvantage | 8.05% (0.70) | 16.59% (0.49) | OR=2.27, 95% CI=1.87-2.77 | OR=1.35, 95% CI=1.10-1.67 |
| Low birth weight | 4.32% (0.52) | 4.10% (0.25) | OR=0.95, 95% CI=0.72-1.25 | - |
| Preterm birth | 4.01% (0.50) | 4.47% (0.26) | OR=1.12, 95% CI=0.85-1.48 | - |
| Smoking in pregnancy | 7.63% (0.67) | 16.14% (0.48) | OR=2.33, 95% CI=1.91-2.84 | OR=1.48, 95% CI=1.20-1.82 |
| Maternal depression | 4.78% (0.54) | 7.71% (0.34) | OR=1.66, 95% CI=1.29-2.14 | OR=1.37, 95% CI=1.05-1.77 |
| Maternal age at birth | 30.09 (0.11) | 28.81 (0.06) | OR=0.94, 95% CI=0.93-0.95 | OR=0.95, 95% CI=0.94-0.96 |
| Maternal education | 3.67 (0.03) | 3.14 (0.02) | OR=0.67, 95% CI=0.64-0.71 | OR=0.71, 95% CI=0.67-0.75 |
| Parity | 0.68 (0.02) | 0.80 (0.01) | OR=1.17, 95% CI=1.09-1.24 | OR=1.12, 95% CI=1.05-1.21 |

* Missing data on indicators used to derive weights were singly imputed as the modal or mean value (all <10% missing). ** Enrolled in original ALSPAC sample. IPW = inverse probability weighting.

**Supplementary Table 12.** Associations between variables used to specify the trajectories and missing covariate data

| | SCDC mean (SE) | | Association with missingness |
|---|---|---|---|
| | Complete covariate data | Incomplete covariate data | (Univariable association) |
| Age 7 SCDC | 2.33 (0.08) | 2.91 (0.05) | OR=1.05, 95% CI=1.03-1.07 |
| Age 10 SCDC | 1.87 (0.07) | 2.50 (0.05) | OR=1.06, 95% CI=1.04-1.08 |
| Age 13 SCDC | 2.04 (0.08) | 2.68 (0.05) | OR=1.06, 95% CI=1.04-1.078 |
| Age 17 SCDC | 2.42 (0.09) | 2.99 (0.06) | OR=1.04, 95% CI=1.03-1.06 |
| Age 25 SCDC | 1.22 (0.07) | 1.81 (0.07) | OR=1.06, 95% CI=1.04-1.09 |

SCDC = Social Communication Disorders Checklist

**Supplementary Table 13.** Associations between covariates and missing covariate data

| | Proportion with associated feature ('covariate') (SE) | | Association with missingness |
|---|---|---|---|
| | Complete covariate data | Incomplete covariate data | (Univariable association) |
| Childhood ASD diagnosis | 0.44% (0.17) | 0.63% (0.10) | OR=1.43, 95% CI=0.64-3.19 |
| High-risk for childhood ASD | 6.57% (0.62) | 9.85% (0.37) | OR=1.55, 95% CI=1.25-1.93 |
| High-risk for adult ASD: parent-rated | 5.88% (0.59) | 6.93% (0.48) | OR=1.19, 95% CI=0.92-1.54 |
| High-risk for adult ASD: self-rated | 13.84% (0.87) | 15.55% (0.84) | OR=1.15, 95% CI=0.95-1.39 |
| Low childhood IQ | 2.97% (0.43) | 7.21% (0.38) | OR=2.54, 95% CI=1.86-3.46 |
| Child pragmatic language problems | 1.64% (0.32) | 3.33% (0.24) | OR=2.06, 95% CI=1.36-3.12 |
| Adult communication problems | 5.69% (0.58) | 9.07% (0.56) | OR=1.65, 95% CI=1.29-2.12 |
| Childhood peer problems | 5.69% (0.58) | 6.73% (0.34) | OR=1.20, 95% CI=0.94-1.52 |
| Adolescent peer problems | 6.64% (0.63) | 7.75% (0.43) | OR=1.18, 95% CI=0.94-1.49 |
| Adult peer problems: parent-rated | 6.70% (0.63) | 8.61% (0.54) | OR=1.31, 95% CI=1.03-1.66 |
| Adult peer problems: self-rated | 14.03% (0.87) | 17.69% (0.84) | OR=1.32, 95% CI=1.10-1.58 |
| NEET | 3.79% (0.48) | 5.02% (0.51) | OR=1.34, 95% CI=0.96-1.87 |
| Distress and impairment: parent-rated | 4.17% (0.50) | 5.58% (0.45) | OR=1.36, 95% CI=1.01-1.83 |
| Distress and impairment: self-rated | 7.90% (0.68) | 10.85% (0.69) | OR=1.42, 95% CI=1.13-1.79 |

NEET = Not in Education, Employment or Training

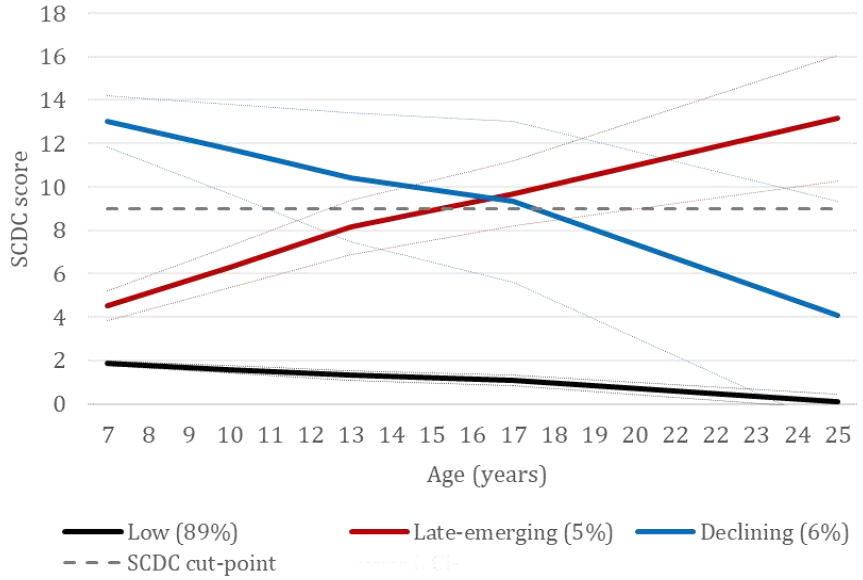
**Supplementary Table 14.** Associations between variables include in the inverse probability weights and exclusion from primary sample

| | Exposure proportion (%) or mean (SE) | | Association with missingness | |
|---|---|---|---|---|
| | In primary sample | Not in primary sample | Univariable association | Multivariable associations from IPW model* |
| Original enrolment** | 96.19% (0.21) | 90.97% (0.35) | OR=1.94, 95% CI=0.35-0.46 | OR=0.29, 95% CI=0.25-0.34 |
| Male sex | 50.23% (0.56) | 52.00% (0.62$ | OR=1.07, 95% CI=1.01-1.15 | OR=1.11, 95% CI=1.04-1.19 |
| Social disadvantage | 14.82% (0.41) | 37.49% (0.68) | OR=3.45, 95% CI=3.16-3.76 | OR=1.77, 95% CI=1.61-1.94 |
| Low birth weight | 4.14% (0.22) | 6.42% (0.32) | OR=1.59, 95% CI=1.36-1.85 | OR=1.39, 95% CI=1.14-1.68 |
| Preterm birth | 4.38% (0.23) | 5.88% (0.31) | OR=1.37, 95% CI=1.17-1.59 | OR=1.05, 95% CI=0.87-1.28 |
| Maternal depression | 7.10% (0.30) | 12.02% (0.47) | OR=1.79, 95% CI=1.58-2.02 | OR=1.18, 95% CI=1.03-1.34 |
| Maternal age at birth | 29.07 (0.05) | 26.56 (0.07) | OR=0.90, 95% CI=0.89-0.90 | OR=0.91, 95% CI=0.91-0.92 |
| Maternal education | 3.25 (0.14) | 2.55 (0.02) | OR=0.64, 95% CI=0.62-0.66 | OR=0.75, 95% CI=0.73-0.78 |
| Parity | 0.77 (0.01) | 0.94 (0.02) | OR=1.18, 95% CI=1.14-1.23 | OR=1.15, 95% CI=1.11*1.20 |

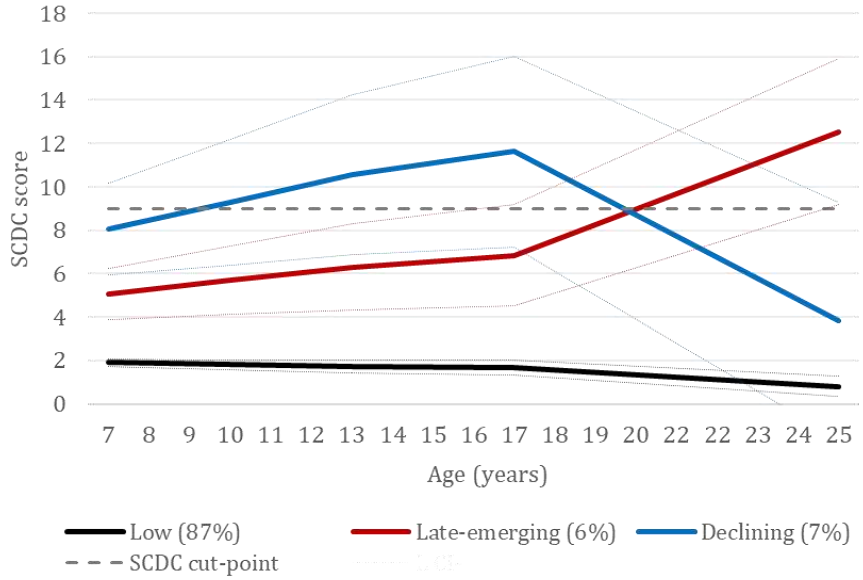
* Missing data on indicators used to derive weights were singly imputed as the modal or mean value where (all <20% missing: smoking in pregnancy excluded as 20% missing). ** Enrolled in original ALSPAC sample. IPW = inverse probability weighting.

**Supplementary Figure 1.** Social Communication Disorders Checklist (SCDC) by class: mean trajectory with 95% confidence intervals

a) Primary sample

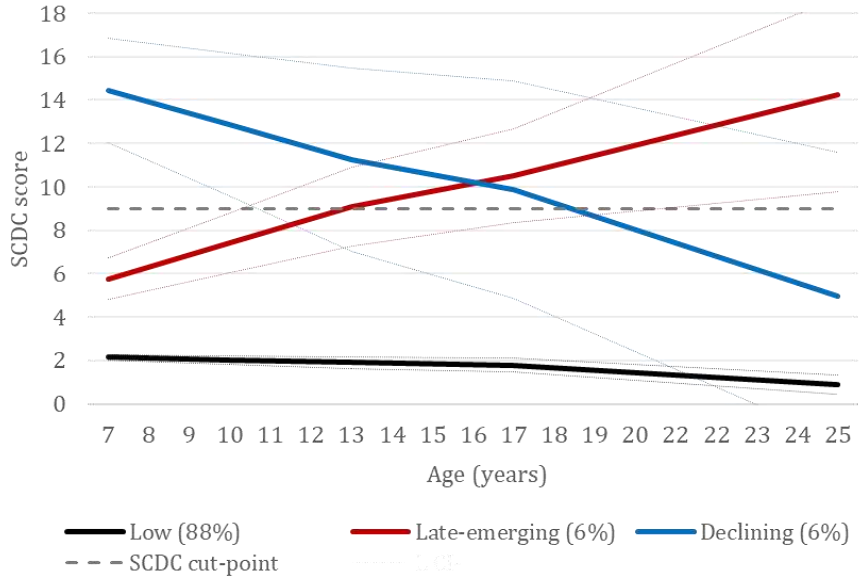



b) Primary sample with regular parent-offspring contact in adulthood

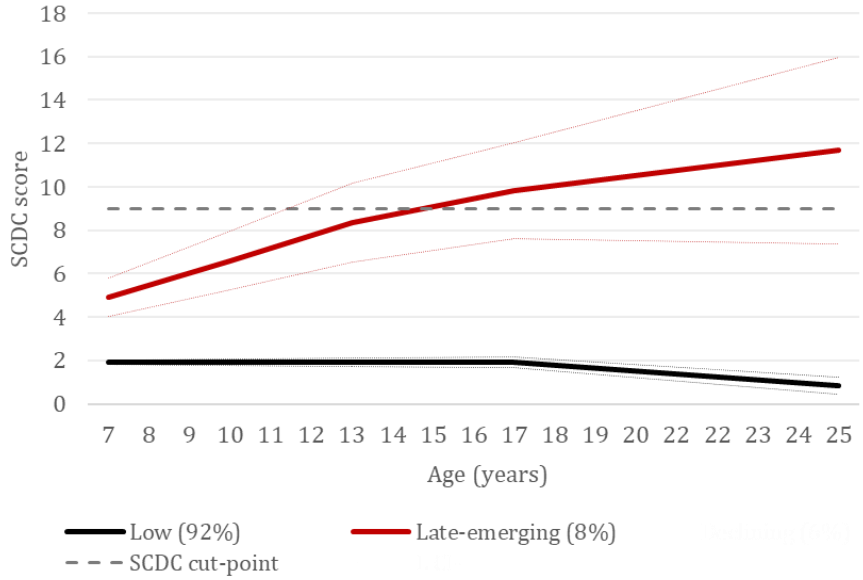


**Supplementary Figure 2.** Social Communication Disorders Checklist (SCDC) by class: mean trajectory with 95% confidence intervals by sex

a) Males

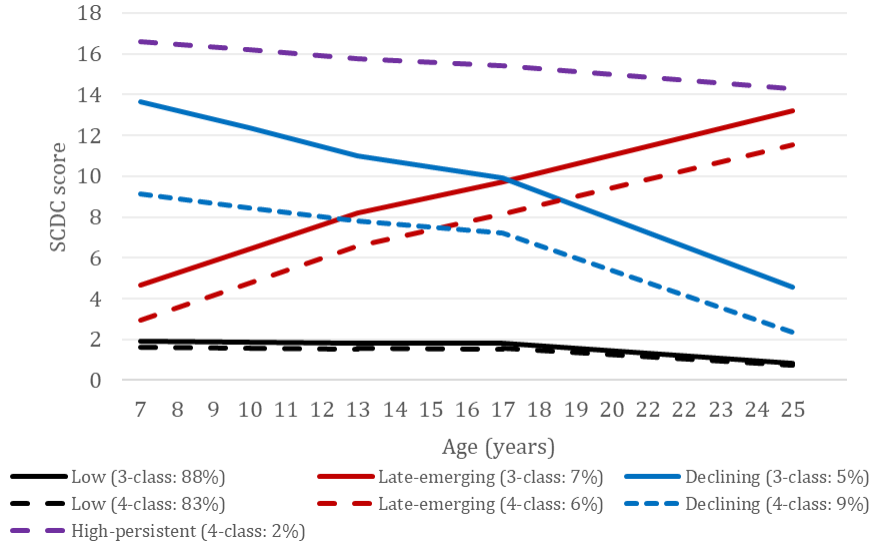



b) Females

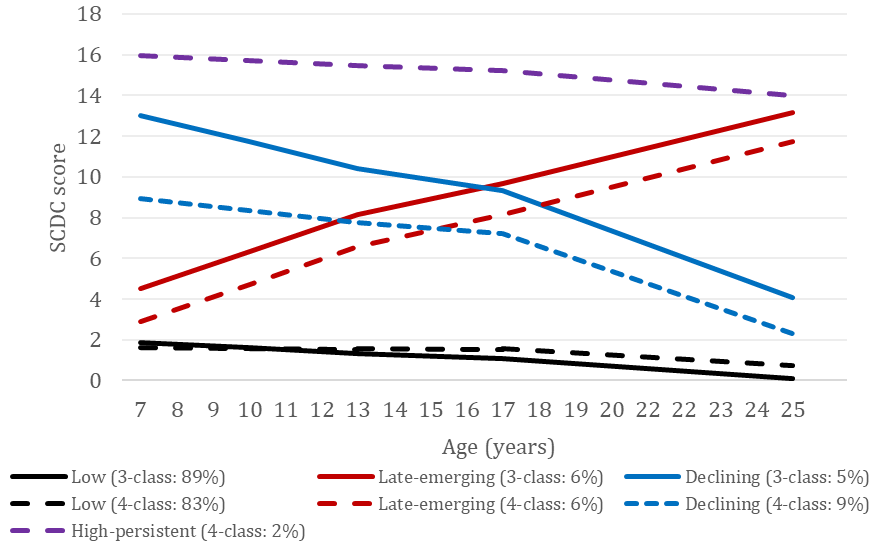


**Supplementary Figure 3a-c.** Social Communication Disorders Checklist (SCDC) by class: 3-class (solid lines) and 4-class (dashed lines) solutions derived with varying levels of missingness

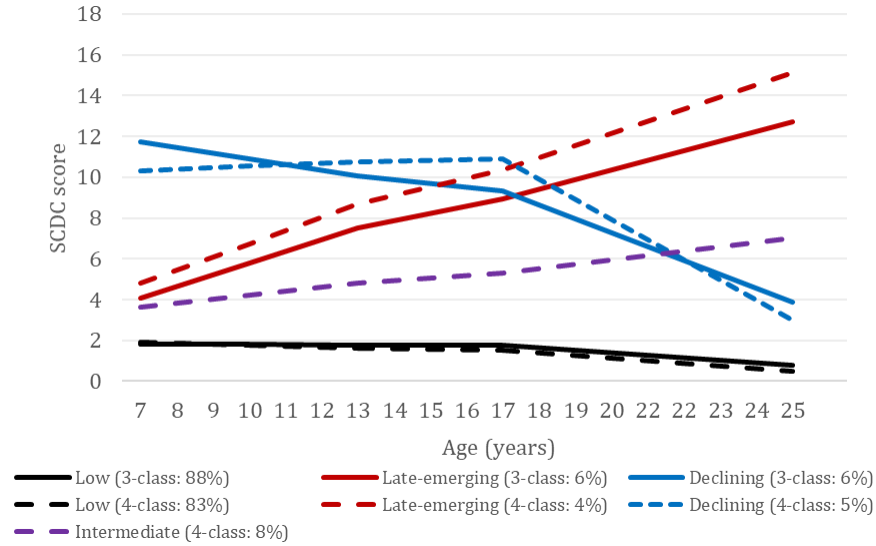
a) 1+ data-points: N=9715




b) 2+ data-points: N=8094 (primary analyses)

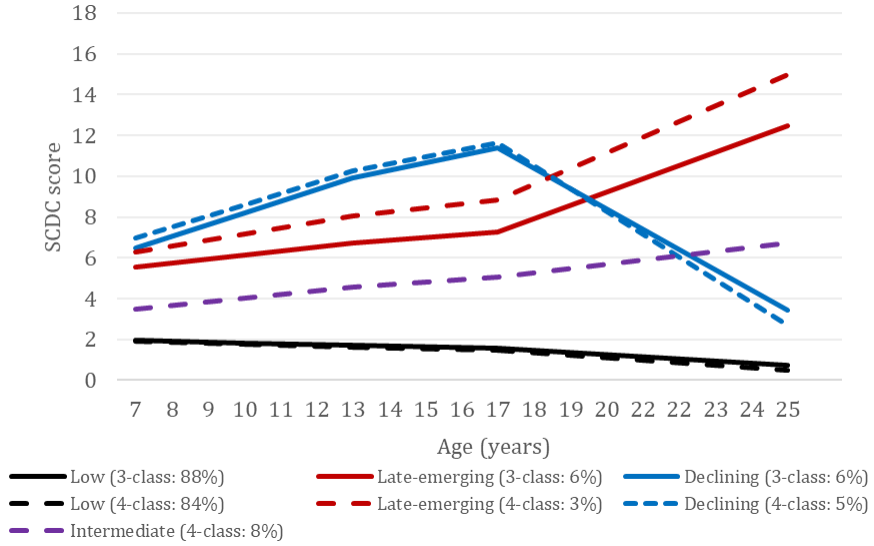



c) 3+ data-points: N=6614

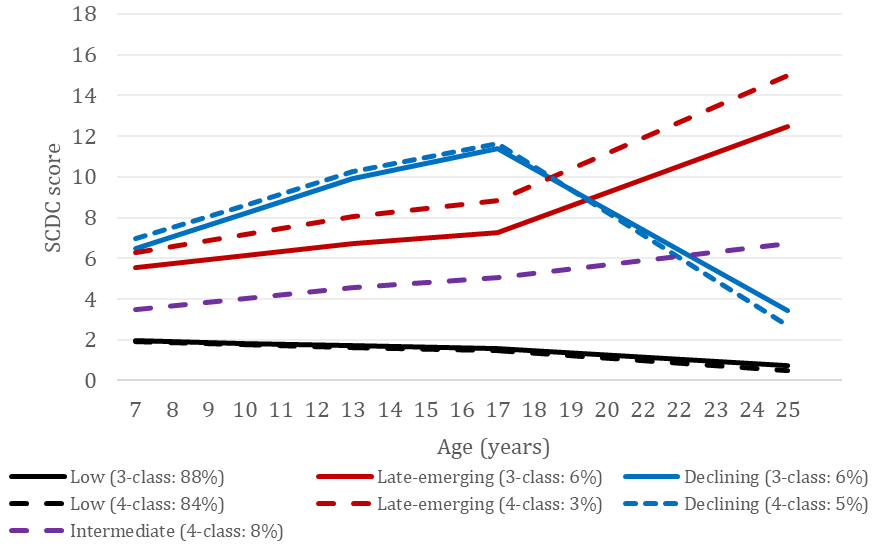


**Supplementary Figure 3d-e.** Social Communication Disorders Checklist (SCDC) by class: 3-class (solid lines) and 4-class (dashed lines) solutions derived with varying levels of missingness

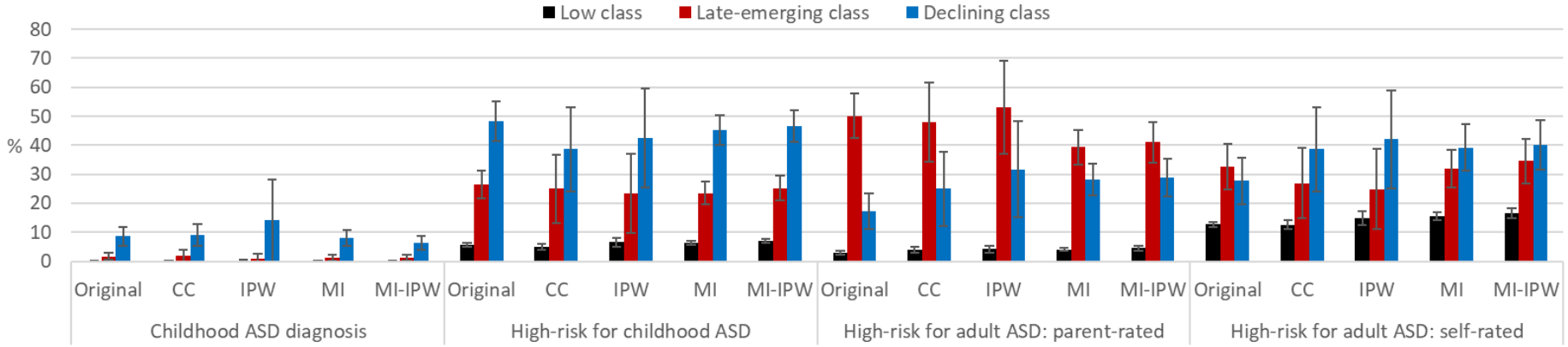
d) 4+ data-points: N=5127




e) 5 data-points: N=3021 (complete cases)

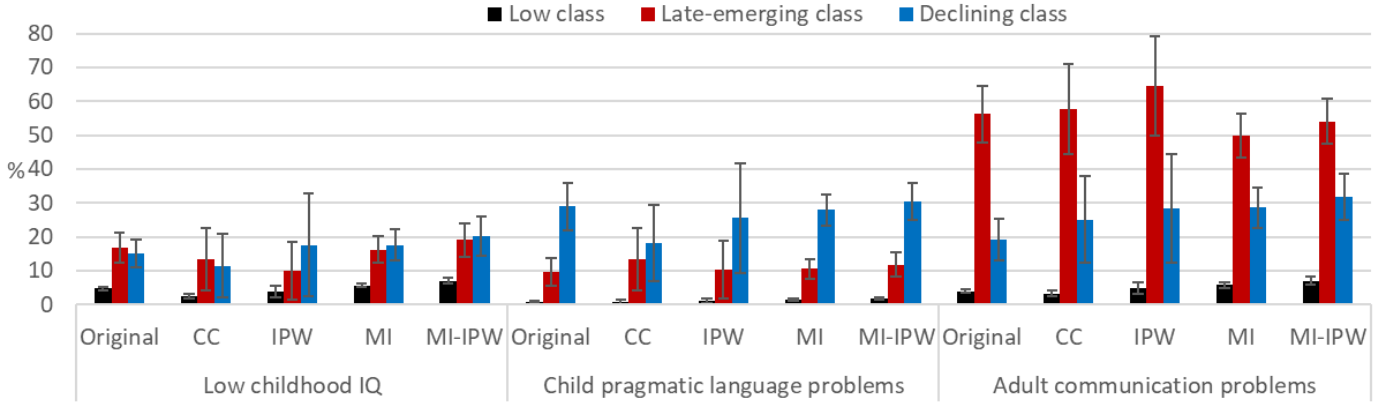


**Supplementary Figure 4.** Prevalence of ASD by trajectory class using different approaches to handle missing data



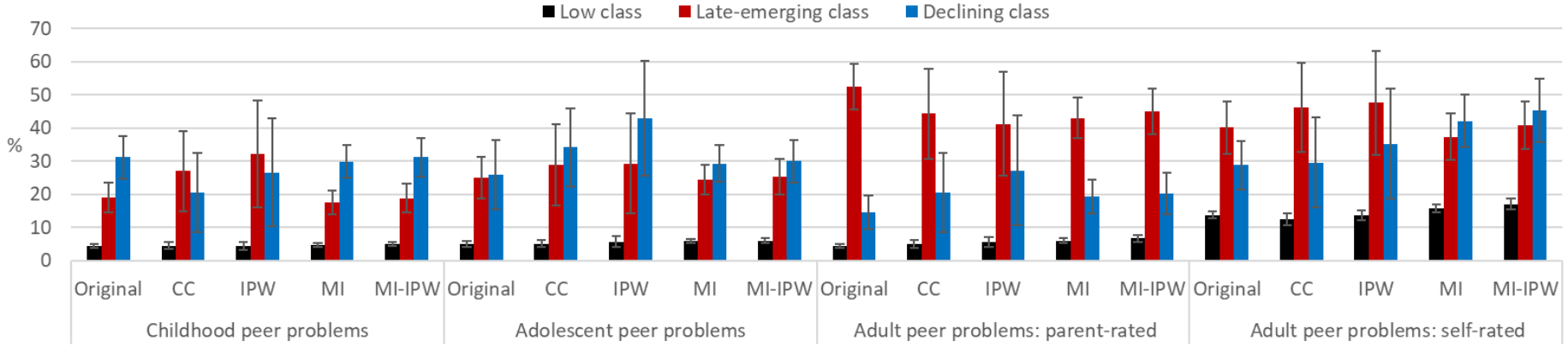

Error bars depict 95% confidence intervals. Original = original estimate, CC = complete cases, IPW = inverse probability weighting, MI = multiple imputation

**Supplementary Figure 5.** Prevalence of low IQ and communication problems by trajectory class using different approaches to handle missing data



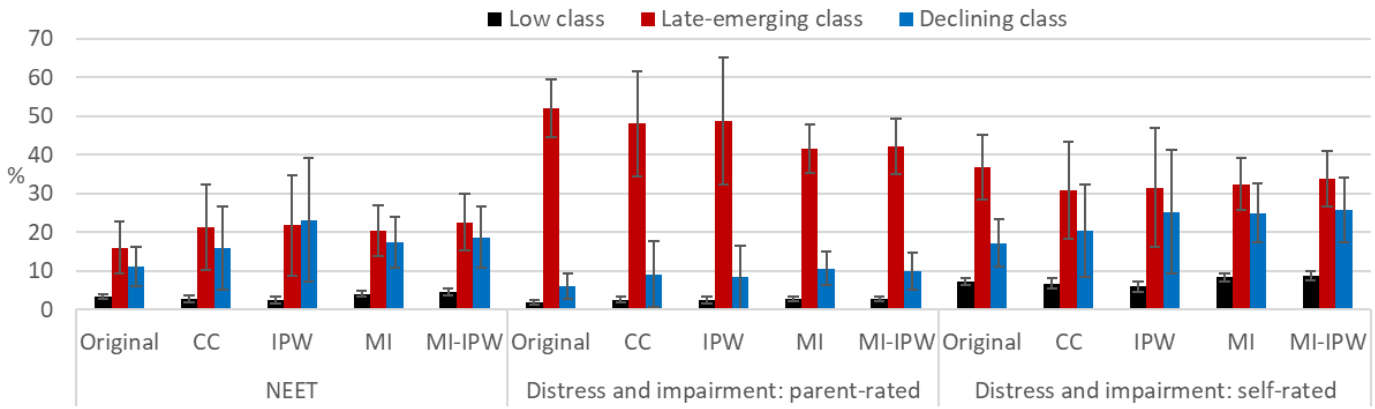
Error bars depict 95% confidence intervals. Original = original estimate, CC = complete cases, IPW = inverse probability weighting, MI = multiple imputation

**Supplementary Figure 6.** Prevalence of peer problems by trajectory class using different approaches to handle missing data




Error bars depict 95% confidence intervals. Original = original estimate, CC = complete cases, IPW = inverse probability weighting, MI = multiple imputation

**Supplementary Figure 7.** Prevalence of impaired adult functioning by trajectory class using different approaches to handle missing data




Error bars depict 95% confidence intervals. Original = original estimate, CC = complete cases, IPW = inverse probability weighting, MI = multiple imputation

## References


1. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. International journal of epidemiology. 2013;42(1):111-27.
2. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International journal of epidemiology. 2013;42(1):97-110.
3. Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson N, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. Wellcome Open Res. 2019;4:51-.
4. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental review. 2016;41:71-90.
5. Muthén LK, Muthén BO. Mplus User's Guide. Seventh ed. Los Angeles, CA: Muthén & Muthén; 1998-2012.
6. Hu LT, Bentler PM. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. Struct Equ Modeling. 1999;6(1):1-55.
7. Cheung GW, Rensvold RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal. 2002;9(2):233-55.
8. Watson B. Young people not in education, employment or training (NEET), UK: February 2020: estimates of young people (aged 16 to 24 years) who are not in education, employment or training, by age and sex. Office for National Statistics. 2020.
9. 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. International journal of epidemiology. 2018;47(4):1207-16.
10. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res. 2013;22(3):278-95.
11. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99.
12. von Hippel PT. How Many Imputations Do You Need? A Two-stage Calculation Using a Quadratic Rule. Sociol Method Res. 2018:doi: 10.1177/0049124117747303.
13. Seaman SR, White IR, Copas AJ, Li L. Combining multiple imputation and inverse-probability weighting. Biometrics. 2012;68(1):129-37.
14. Hoekstra RA, Vinkhuyzen AA, Wheelwright S, Bartels M, Boomsma DI, Baron-Cohen S, et al. The construction and validation of an abridged version of the autism-spectrum quotient (AQ-Short). J Autism Dev Disord. 2011;41(5):589-96.