Section 1. Introduction to Multilevel Growth Curve Models in Analyses

The data collected represented repeated caregiver reports of children’s externalizing problems, assessed using the Strengths and Difficulties Questionnaire (SDQ). Therefore, the outcomes of interest were repeated measures at different waves (measurement occasions), nested within individual participants.

Growth models are ideal tools for analysing repeated measures since they provide answers regarding how an outcome changes over time, the inter-individual variability in these changes, and the effects of time-invariant and time-varying covariates. Growth curve models describe patterns of change over time by estimating parameters that represent the average initial value of the outcome at the point chosen as the start of the data collection, the *intercept*, and the average rate of change for the time unit, the *slope*. The point representing the start of data collection may be changed to choose a more meaningful one: for example, studies of children development may choose birth as the conventional starting point. In our study, considering an intercept that represented average levels of SDQ problem behaviour scores at birth would be meaningless, we chose a conventional starting point at a later age (see below). The time unit for describing changes in the outcome can also be decided according to the nature of the outcome observed and to the study design (e.g., months or years).

Growth models can be parameterized as latent variable models (whereby the intercept and slope parameters are considered latent variables explaining the observed outcomes), or as multilevel regression models. We opted for multilevel regression models as they can be more easily adapted to more complex models that allow for non-linear patterns of change and take into account differences in the timing of data collection (Hox et al., 2005; Macdonald-Wallis et al. 2012). A comprehensive guide to multilevel growth curve models is provided by the first part of the book by Singer and Willett (2003) referenced below.

The time unit in the study was children’s chronological age in years. Children’s age in years was calculated in each completed wave of data collection. We centred this at age 3 years, so that the intercept represented the estimated sample average SDQ problem scores at age 3 years. The simplest growth curve model assumes a linear rate of change over time. This model is formally specified as:

Where:

represents the expected problem score *y* for child *i* at time *j*;

represents the age of child *i* at time *j;*

represents the intercept for child *i* , that is, the expected problems score *y* for child *i* when , i.e. 3 years;

represents the slope for child *i*, that is, the expected change in problems score for a 1-unit difference in age;

represents deviation from the expected problems score at time *j* forchild *i*, which is supposed to be approximately normally distributed with mean 0 and variance :

(S1.4) ) ;

represents the sample average of the children’s intercepts;

represents the sample average of the children’s slopes;

represents the amount of deviation from the expected sample average intercept for child *i*;

represents the amount of deviation from the expected sample average slope .

The latter two parameters are supposed to approximate a normal bivariate distribution with averages equal to zero and variances as follows:

Where:

represents the sample residual variance of the estimated intercept;

represents the sample residual variance of the estimated slope;

represent the residual covariance between the intercepts and slopes across all children in the sample ( ).

Equations S1.1, S1.2, and S1.3 can be used to determine a composite equation:

;

where all terms have the same interpretation as provided above. The terms labelled by Greek letter *γ (gamma)* are the *fixed* effects in the equation, insofar they represent average sample parameters, whereas the *u* and *e* terms represent random effects: namely, the *u* terms represent inter-individual variability around the fixed effects, while the *e* effect represents intra-individual variability across time.

Equation S1.6 facilitates the inclusion of covariates. For example, considering a time-invariant covariate, e.g., sex at birth, coded as male, we can assume that difference in sex may be associated with differences in the average level of problem behaviour *y*:

;

In equation S1.7 the parameter represents the sample average difference in the intercepts if male=1, i.e. for male participants.

In a similar manner, it is also possible to assume that covariates affect the rate of change of the outcome:

+ ] +

;

In equation S1.8 the added parameter represents the sample average difference in change over time when male = 1, i.e. how the rate of change of the outcome changes for male participants (as opposed to females).

More complex patterns of change over time can be modelled by adding polynomials of time. For example, adding a quadratic term to equation S1.6 would indicate the rate of change of the outcome can accelerate or decelerate:

;

The parameter represents the average sample rate of change associated with a 1-unit change in age squared, and represents child *i*‘s deviation from the average quadratic slope. The sign of indicates whether the rate of change in the outcome accelerates over time (or decelerates over time (

Models with additional time polynomials are nested, and therefore their fit can be compared using Likelihood Ratio tests. In our analyses, we tested linear, quadratic, and cubic polynomials for each of the two main outcomes (SDQ Internalizing and Externalizing composite scores). Models can also be simplified by removing parameters that concern covariances or variances. In our analyses we tested the removal of covariances and variances (e.g., removal of individual variance around the quadratic slope, see S1.9) using Likelihood Ratio tests to assess whether removal of these parameters was warranted.

We estimated models using Stata 17 *mixed* command (StataCorp, 2021). The command uses Maximum Likelihood methods by default. These methods allow the estimation of parameters across participants that have complete or incomplete outcome data. This estimation method has been proved to be more reliable than less sophisticated methods that deal with missing data in repeated measure studies (e.g., mean imputation), as long as data are missing at random. The assumption of missing at random implies that while some covariates may explain the pattern of missing outcome data (e.g., missing data are more likely in families from more deprived background), the pattern of missing outcome data is not related to the unobserved outcomes themselves. We considered the assumption of missing at random plausible in our study, thus reporting parameters estimated across all participants that had provided data on at least one measurement occasion (data collection wave). The number of children that had only one or two SDQ questionnaires completed across the study was negligible, see Table S1.1: therefore, we did not exclude these participants from analyses of growth curves. However, we also run sensitivity analyses to test whether we observed the same pattern of results when we excluded participants who completed fewer than three waves of data collection. Missing data were also present in some covariates for a few participants (see Supplementary Material Section 3). Multilevel growth curve models do not impute missing data for covariates; therefore we adopted multiple imputation to ensure the results included the full sample. The multiple imputation procedure we used is described in the next section of Supplementary Material.

*Table S1.1*

*Number and Percentages of Participants by Number of Data Collection Waves Completed*

|  |  |  |
| --- | --- | --- |
| Number of Data Collection Waves Completed | N | % |
| 1 | 4 | *4.35* |
| 2 | 9 | *9.78* |
| 3 | 15 | *16.30* |
| 4 | 8 | *8.10* |
| 5 | 28 | *30.43* |
| 6 | 28 | *30.43* |
| Total | 92 | *100.00* |

References

Hox, J., & Stoel, R. (2005). Multilevel and SEM Approaches to Growth Curve Modeling, 1296-1305. In: Brian Everitt & David Howell (Eds.) *Encyclopedia of Statistics in Behavioral Science*. New York: Wiley and Sons. <https://doi.org/10.1002/9781118445112.STAT06603>.

Macdonald-Wallis, C., Lawlor, D., Palmer, T., & Tilling, K. (2012). Multivariate multilevel spline models for parallel growth processes: application to weight and mean arterial pressure in pregnancy. *Statistics in Medicine*, *31*, 3147 - 3164. <https://doi.org/10.1002/sim.5385>.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford: Oxford University Press. [https://doi.org/10.1093/acprof:oso/9780195152968.001.0001](https://psycnet.apa.org/doi/10.1093/acprof:oso/9780195152968.001.0001)

StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC.

Section 2. Multilevel Growth Curve Models in Analyses

All *N* = 92 participants in the sample had complete data on covariates sex, SES component score of adoptive family, child’s age at time of placement. However, a total of 83 (90% of all participants) and 78 (85% of all participants) had valid information on the total of ACES they had been exposed to and the number of moves before placement. Since these were considered important covariates to include in the analyses and we wanted to ensure results of the main multilevel growth curves included all 92 participants, thus minimising bias, we used multiple imputation (MI) to estimate plausible values of the missing covariates in those cases where the covariate was missing.

The purpose of multiple imputation (MI) is to create series of datasets whereby the missing values in one or more variables are replaced by plausible imputed values. Statistical analyses and models are run on the imputed datasets, and the parameter estimates averaged across the *n* imputed datasets. Standard errors are calculated taking into account variability across the *n* imputed datasets, thus adequately representing uncertainty associated with the missing values.

Multiple imputation is considered a valid method for tackling missing data as long as the missing values can be plausibly assumed to be Missing Completely at Random (MCAR) or Missing at Random (MAR). MCAR describes a scenario whereby there are no systematic differences between the missing and the observed values. MAR describes a scenario whereby the differences between missing and observed values are associated with differences in other observed variables. A different scenario whereby Multiple Imputation would not be warranted involves data Missing not at Random (MNAR): in this case, the reason for data being missing is systematically associated with values of the missing variable itself. We assumed that the mechanisms underlying missing information in our datasets were most likely to be of the MAR type.

Because our data had missing values on two covariates, we used chained equation methods to create 250 imputed datasets. Chained equations are used to fill missing data in an iterative manner. We considered categories of these covariates, hence, in order to obtain categorical imputed values, the imputation of missing values was carried out using ordered logistic regressions. To impute plausible missing values, we added covariates in the imputation process such as: the average internalizing problem scores; the average externalizing problem scores, age at time of placement, and the family SES component score.

We conducted multiple imputation in Stata software version 17.0 (StataCorp., 2021), using the ‘mi impute chained’ command suite. After creating the 250 imputed datasets for all cases in the analyses, parameters were estimated using ‘mi estimate:’ command in Stata. These methods use Rubin’s rule to average parameter estimates across the datasets with imputed complete values.

**References**

StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC.

Section 3. Detailed Results of Multilevel Growth Curve Models

**SDQ Internalizing**

An initial model with a single linear slope indicated a significant fit to the data, Wald *χ2*(1) =28.98, *p* < .001, with the coefficient of the linear slope being significant, *coefficient* = 0.37, *z* = 5.38, *p* < .001. The model with a quadratic slope, as well as a quadratic slope variance and covariances (see Equation S1.9 in Supplementary Material section 1) did not converge. Inclusion of a quadratic slope (but excluding the quadratic slope variance) did not demonstrate a significant increase in model fit, LR *χ2*(1) =0.02, *p* = .89, and neither did the inclusion of cubic polynomial of age, LR *χ2*(2) =0.61, *p* = .74.

Further model checks also indicated that removal of the covariance between intercept and slope did not significantly decrease model fit, LR *χ2*(1) =0.68, *p* = .41, but removal of the slope variance did result in a worsening of model fit, LR *χ2*(1) =36.67, *p* < .001: thus, we retained the slope variance parameter (see parameters reported in the main manuscript for more information).

Further model checks investigated if covariates sex (dummy coded as male), family SES, and age at placement had an effect on the initial status and the slope of the outcome. Including an interaction term representing the interaction between age and male (see Equation S1.8 in Supplementary Material 1) did not indicate a significant improvement in model fit, LR *χ2*(1) =0.77, *p* = .38. The results suggest there were not significant differences in the rate of change of internalizing problem scores between males and females. Similarly, the inclusion of an interaction between family SES and age did not indicate a significant improvement in model fit, LR *χ2*(1) =0.58, *p* = .45. Inclusion of age at time of placement indicated that the average internalizing problem scores changed in accordance with the child’s age at placement, but the results did not indicate that the internalizing problem scores rate of change varied according to the age of placement, LR *χ2*(4) =3.17, *p* = .53.

The final model with covariates is reported in the main manuscript. This model was estimated using multiple imputation of the categorical covariates “number of ACEs” and “number of moves”, as detailed in the previous Supplementary Material Section 2.

**SDQ Externalizing**

An initial model with a single linear slope indicated a significant fit to the data, Wald *χ2*(1) =25.39, *p* < .001, with the coefficient of the linear slope being significant, *coefficient* = 0.34, *z* = 5.04, *p* < .001. Inclusion of a quadratic slope (see Equation S1.9 in Supplementary Material section 1) demonstrated a significant increase in model fit, LR *χ2*(2) =29.23, *p* < .001. Inclusion of a cubic polynomial of age did not indicate a significant improvement in model fit, LR *χ2*(1) = 0.06, *p* = .81. Thus, the quadratic model was retained.

Further model checks indicated that removal of the covariances of the quadratic slopes resulted in a significant decrease in model fit, LR *χ2*(3) = 20.01, *p* = .0002, and so did the removal of the variance of the quadratic slope (and relative covariances) LR *χ2*(3) =19.78, *p* = .0002. Thus, these variances and covariances were retained in the final model: parameters estimates are reported in the main manuscript.

Further model checks investigated if covariates sex (dummy coded as male), family SES, and age at placement had an effect on the initial status and the slope of the outcome. Including an interaction term representing the interaction between age and male (see Equation S1.8 in Supplementary Material 1) did not indicate a significant improvement in model fit, LR *χ2*(1) =1.71, *p* = .43. The results suggest there were not significant differences in the rate of change of externalizing problem scores between males and females. Similarly, the inclusion of an interaction between family SES and age did not indicate a significant improvement in model fit, LR *χ2*(1) =0.02, *p* = .02. Inclusion of age at time of placement did not indicate that the externalizing problem rate of change varied according to the age of placement, LR *χ2*(6) =6.60, *p* = .36.

The final model with covariates is reported in the main manuscript. This model was estimated using multiple imputation of the categorical covariates “number of ACEs” and “number of moves”, as detailed in the previous Supplementary Material section 2.

**Sensitivity Analyses**

The analyses reported in the manuscript and the previous two sections included *N* = 92 participants who had incomplete data in the outcomes of interest. In particular, *n* = 4 and *n =* 9 participants had only completed one and two waves of data collection respectively. However, since we were modelling trajectories of the outcomes over periods that spanned six measurement occasions, the appropriateness of including participants with fewer than three completed waves of data collection in the estimation of trajectories may be questioned (e.g., Twisk, 2013). For this reason, we ran sensitivity analyses where the trajectories of the two outcomes were estimated on the sub-sample that had completed three or more waves of data collection (*n* = 79). The pattern of results was consistent with the results reported in the main manuscript.

When we considered the internalizing scores, the results indicated a best-fitting model akin to the model estimated on the *N =* 92 sample: the model estimated for the *n* = 79 subsample indicated significant variation around the sample intercept and time slope, and a linear coefficient with an estimated 0.35 score increase per year, *95% CI*  0.11 to 0.31, *z =* 5.07 ,  *p* < .001.

When we considered externalizing scores, the best fitting model for the *n* = 79 subsample was also very similar to the model estimated on the *N =* 92 sample. The best fitting model indicated a quadratic change over time with the linear slope estimated as 0.68 (95% *CI* 0.36 to 1.01, *z* = 4.09, *p* < .001) and the quadratic slope estimated as -0.04 (95% *CI* -0.07 to -0.01, *z* = -2.68, *p* = .007). The model included significant variation around the intercept and slopes, as well as covariances between these random parameters.

Overall, these sensitivity analyses indicated that when we considered participants who had provided data on three waves or more, the models and their parameters were very similar to those based on the *N =* 92 sample. Detailed results of these analyses are not reported but can be provided if requested.

**References**

Twisk, J.W.R. (2013). Missing data in longitudinal studies. In: Twisk, J.W.R., *Applied longitudinal data analysis for epidemiology*. Cambridge: Cambridge University Press <https://doi.org/10.1017/CBO9781139342834>

Section 4. Introduction to Spline Models used in the analyses

Splines are models for repeated measurements that specify different polynomial functions for different intervals of time. This allows for greater flexibility in modelling behaviour change over time. More common models (e.g. linear growth models) assume that change follows the same function over time (e.g. a linear increase or decrease in the outcome of interest, see the first Supplementary Material section). Spline models instead can specify different functions in different periods of time (e.g. a linear increase followed by a linear decrease), thus representing patterns of change that vary across time. An example where spline models are particularly apposite is the study of children growth in different countries (see Howe et al., 2013).

Another key advantage of spline models lies in providing a single continuous function that applies to the whole study period. Namely, spline models allow modelling complex patterns of change over time using a relatively simple function. Spline models achieve this by dividing the study period into intervals with overlapping fixed points called *knot points*. Within these intervals, the polynomial functions influence on the outcome is controlled by weights that represent the passing of time within the interval. The knot points are thus the points where two polynomial functions meet, allowing the polynomials to model behaviour change with no discontinuity from one interval to another. In other words, the knot points act as “pivots” for the polynomials that describe behaviour change over time (McElreath, 2020).

To summarise, spline models are random slope models akin to models where the behaviour of interest changes as a function of time between measurement occasions. However, the key feature of spline models is that rather than considering time as a single explanatory variable, time *t* is divided into smaller intervals *s* (for *splines*) determined by the definition of knot points *c*. Each of these *s* intervals have a different slope (or more slopes to allow for polynomials): these slopes represent different directions and rates of change within the specific intervals.

More formally, a spline model is described as:

Where:

indicates the predicted value of outcome *y* at occasion *j* for child *i*;

indicates the sample average value of outcome *y* when all the splines are set to 0, which is usually the start of the study, the average of individual intercepts;

indicates child *i* deviation from the sample average score when all splines are equal to 0;

indicates the sample average rate of change (the slope) during interval *k*;

indicates child *i* deviation from the average rate of change within interval *k*;

indicates the a child *i* measurement occasion *j*  within interval *k*;

indicates deviation from the predicted score of individual *i* at time *j,* and it is supposed to approximate a normal distribution with mean 0 and variance :

(S4.2) )

Additionally, the random terms and are supposed to follow a bivariate normal distribution with means 0 and covariance:

(S4.3)

The knot points are set at the start and end of the observation period: within these two extremes, researchers can specify different numbers of knot points to create *k* splines *s* as follows (the first column describes the value of the spline, the second column describes the condition that determines that value):

(S4.4)

A key issue in the application of spline models is the definition of the knot points, i.e. the points where a study period is split into different intervals (see Howth et al., 2016). If, for example, two knot points are selected, three time intervals are created, i.e. three splines *s* with different slopes, see equation (S4.1). Thus, the number of parameters can escalate. The choice of knot points may be somewhat arbitrary, and therefore it requires justification and rationale. In the absence of a strong rationale, sensitivity analyses may be required to test the consistency of results to different knot points specifications.

In the analyses we reported, we selected a single knot point based on our substantive interest in modelling children’s pattern of change *before* and *after* a key transition point, that is, children’s entry into primary school. The knot point was thus the exact age when a participating child started school. School entry takes place at the start of the curriculum year in September of each calendar year in England and Wales: The average age at the time of school entry for the whole *N*=96 sample originally recruited was 4.50 years (*SD* = 0.28), with range between 4.00 and 4.99 years. To facilitate interpretation of results, participants’ age was centred at 3 years: namely, age was expressed as time (in years) from age 3 years.

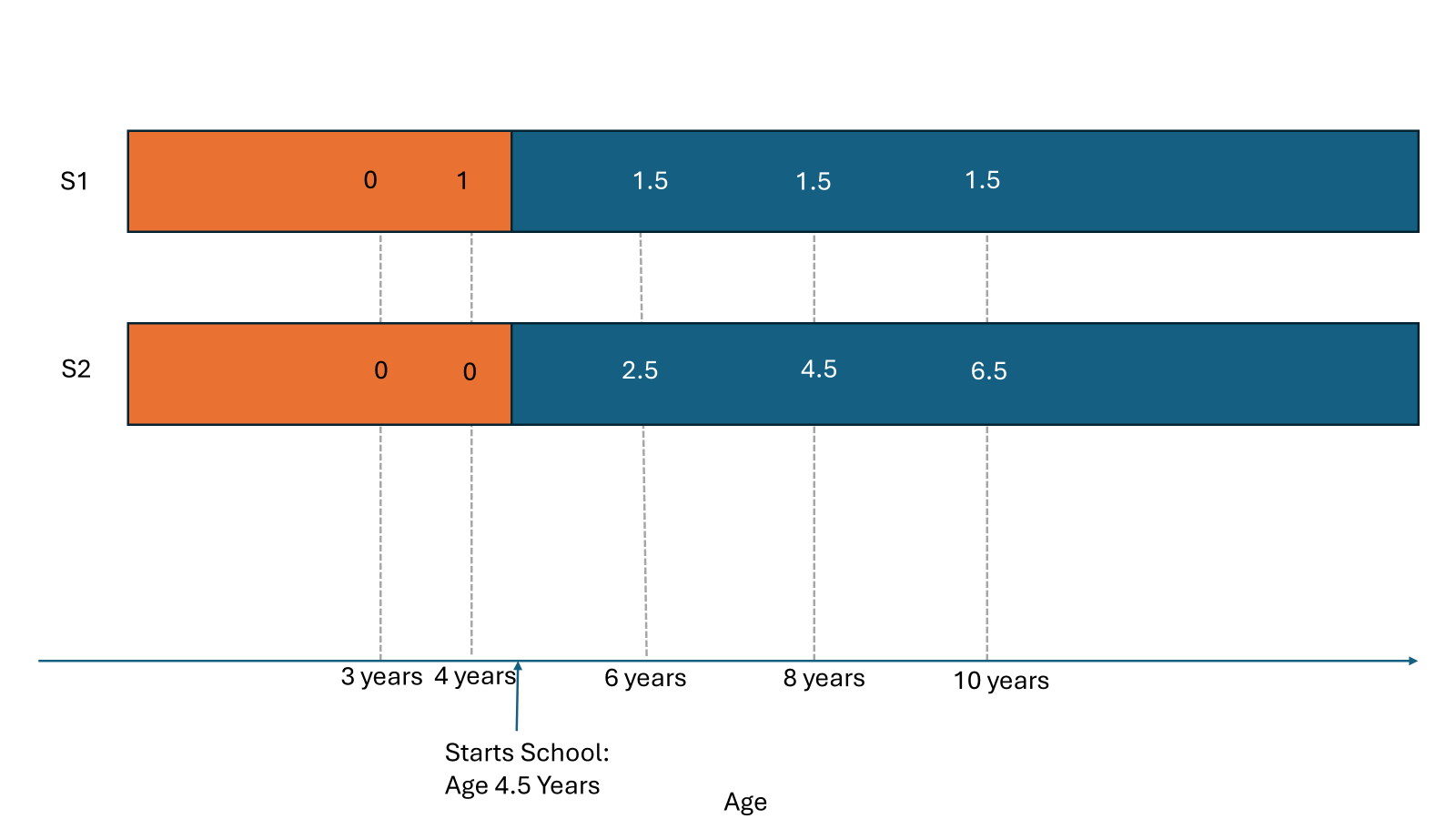
The definition of a single knot point created two spline variables where, if we define as the centred age of participant *i* at measurement occasion *j*, and the (centred) age when participant started *i* school, the two splines values are:

(S4.5)

(S4.6)

Figure 1 below illustrates the specification of the two splines for a hypothetical participant who started school at 4.5 years.

**Figure 1**

*Representation of Spline Values for a Hypothetical Participant*

*Note. The individual us supposed to have been observed on five occasions at 3, 4, 6, 8, and 10 years of age, respectively.*

The values in spline 1 (S1) is equivalent to the participant age, centred at 3 years. Once the individual enters school, i.e. reaches the knot point, the values of S1 remain fixed at the individual’s (centred) age at the knot point. The values of spline 2 (S2) are 0 before the individual reaches the knot point (school entry) and are equivalent to the number of years after the individual passes the knot point.

Since these models prescribed comparisons between the SDQ scores before and after school entry, only *n* = 62 participants that had at least one measurement before and after school entry were included in the analyses. These represented approximately 67% of the *N* = 92 sample.

For both outcomes of interest, we initially tested linear age effects in each spline while modelling residual variances around the average intercept and slopes and respective covariances (see Equation S4.3). After this initial step, we tested models where covariance terms were removed. We then tested models with quadratic rate of change in the second spline (S2). Models were compared using Likelihood Ratio (LR) tests. The best fitting models were retained and used to represent predicted scores reported in the text, before adding covariates to the best-fitting model. More detailed results are reported in the following Supplementary Material section.

**References**

Howe, L. D., Tilling, K., Matijasevich, A., et al. (2013) Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research, 25*(5), 1854-1874. https://doi.org/10.1177/0962280213503925

Section 5. Detailed Spline Model Results

In this section we report detailed results of the spline models for the two main outcomes (SDQ internalizing problems and SDQ externalizing problems). A formal description of the spline models is reported in the previous section of the Supplementary Material.

SDQ Internalizing

The first model we tested included linear slopes for the two splines. The model indicated a non-significant linear slope in S1, *coefficient* = -0.52, *z* = -0.22, *p* = .82, and a significant linear slope in S2, *coefficient*= 0.49, *z* = 4.65, *p* < .001. However, the model demonstrated issues in convergence where the standard errors of the variance and covariance terms could not be estimated.

We then estimated a model with linear slopes in the two splines removing the estimation of covariances. This model converged and showed adequate fit to the data, Wald *χ2 (2)* = 22.68, *p* < .0001. The results indicated a non-significant slope in S1, *coefficient* = -0.02, *z* = -0.06, *p* = .95, and a significant slope in S2, *coefficient* = 0.47, *z* = 4.49, *p* < .001. A further model was tested whereby variance around the slope in S1 was removed, but a LR test indicated the removal of this variance significantly decreased model fit, LR *χ2 (1)* = 5.71, *p* = .017, and was therefore rejected. This indicated that there was significant inter-individual variation in the rate of change in S1, the interval before school entry.

A further model included a quadratic slope in S2 to test whether the rate of change in S2 could be described by a non-linear rate of change. The model with the quadratic slope in S2 provided improved fit compared with the model with a linear slope, LR *χ2 (1)* = 9.79, *p* = .001, and was therefore retained and considered the final model. The model demonstrated significant fit to the data, Wald *χ2 (3)* = 26.43, *p* < .0001. Model parameters are reported in Table S5.1 below. The quadratic term indicated a significant increase in internalizing problems scores after school entry, but this increase decelerated slightly with increasing age: Substituting the model parameters to equation (S4.1) in Supplementary Material S4, we can estimate that while internalizing problem scores were expected to increase from 3.11 to 3.46 (a 0.35 points increase) from age 4 to age 5, and were expected to increase from 3.46 to 4.26 (a 0.80 points increase) from age 5 to age 6, this increase decelerated over time: e.g., between 8 and 9 years the expected score increase was just 0.29 points (from 5.35 to 5.64). The trajectory of expected scores based on the model are plotted in the main manuscript.

**Table S5.1**

*Parameters of the Final Spline Model of Internalizing Problems*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effects | Coefficient | *95% CI* LB | *95% CI* UB | z |
| Intercept | 3.33 | 2.76 | 3.90 | 11.47\*\* |
| Linear trend S1 | -0.22 | -0.74 | 0.30 | -0.82 |
| Linear trend S2 | 0.97 | 0.45 | 1.49 | 3.65\*\* |
| Quadratic trend S2 | -0.08 | -0.17 | -0.003 | -2.04\* |
| Random Effects | Coefficient | *95% CI* LB | *95% CI* UB |  |
| Intercept Variance | 2.08 | 0.99 | 4.36 |  |
| Slope S1 Variance | 1.00 | 0.33 | 3.03 |  |
| Slope S2 Variance | 0.23 | 0.10 | 0.55 |  |
| Within-Individuals Variance | 4.66 | 3.72 | 5.85 |  |

*Note*. S1: Spline 1; S2: Spline 2; 95% CI: 95% Confidence Intervals; LB= Lower Bound; UB= Upper Bound. \* *p* < .05 ; \*\* *p* < .001.

Finally, we added covariates sex at birth (dummy coded as male), age at time of placement (categorised) and the standardised SES component score. The results are reported in Table S5.2. Since only 62 children who had been observe before and after school entry were included in the analyses, all these children were in placement before age 4.5, the average time when children started school. While there were no significant associations between the outcome and covariates sex or SES, there was a significant association between internalizing problems and age at time of placement, see Table S5.2. Those that were in placement at a later age and before they started school, were more likely to show elevated internalizing problem scores.

**Table S5.2**

*Parameters of the Final Spline Model of Internalizing Problems with Covariates*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effects | Coefficient | *95% CI* LB | *95% CI* UB | z |
| Intercept | 2.74 | 1.79 | 3.70 | 5.63\*\*\* |
| Linear trend S1 | -0.40 | -0.91 | 0.12 | -1.49 |
| Linear trend S2 | 1.04 | 0.51 | 1.57 | 3.88\*\*\* |
| Quadratic trend S2 | -0.09 | -0.18 | -0.01 | -2.22\* |
| Sex at Birth: Male | 0.22 | -0.75 | 1.99 | 0.45 |
| Age at Placement: Birth | *Reference* |  |  |  |
| Age at Placement: 1 to <3 yrs | 0.58 | -0.46 | 1.61 | 1.09 |
| Age at Placement: 3 to <5 yrs | 2.33 | 0.63 | 4.03 | 2.69\*\* |
| Std SES Component | -0.02 | -0.54 | 0.49 | -0.09 |
| Random Effects | Coefficient | *95% CI* LB | *95% CI* UB |  |
| Intercept Variance | 1.78 | 0.82 | 3.85 |  |
| Slope S1 Variance | 0.77 | 0.21 | 2.87 |  |
| Slope S2 Variance | 0.24 | 0.10 | 0.55 |  |
| Within-Individuals Variance | 4.73 | 3.77 | 5.94 |  |

*Note*. S1 : Spline 1; S2: Spline 2; 95% CI: 95% Confidence Intervals; LB= Lower Bound; UB= Upper Bound. \* *p* < .05 ; \*\* *p* < .01; \*\*\* *p < .001*.

Externalizing Problems

The initial model indicated a significant linear trend in S1, *coefficient* = 0.64, *z* = 2.21, *p* = .027, as well as a significant linear trend in S2, *coefficient* = 0.32, *z* = 3.44, *p* = .001; However, the estimation of variances and co-variances did not return confidence intervals, thus indicating problems in model convergence. A further model removed the estimation of covariance parameters, indicating good fit to the data Wald *χ2(2)* = 24.73, *p* < .0001. The model confirmed a significant linear trend in S1, *coefficient* = 0.65, *z* = 2.15, *p* = .032, and a significant linear trend in S2, *coefficient* = 0.34, *z* = 3.49, *p* < .001.

A further model included a non-linear quadratic slope in S2: the model showed a significant improvement in model fit compared with the model with a linear trend in S2, LR *χ2(1)* = 6.79, *p* = .009. The model with the quadratic slope in S2 showed a significant fit to the data, Wald *χ2(3)* = 32.11, *p* < .0001. Model parameters are reported in Table S5.3. The model indicated a positive but non-significant slope in S1, and a significant quadratic trend in S2.

The quadratic trend indicated a significant increase in externalizing problems after school entry, but this increase decelerated: in contrast to internalizing scores, this deceleration took place at a younger age and the trend indicated expected scores to decrease around age 10, see the figure with predicted scores in the main manuscript. Using equation (S4.1) in Supplementary Material S4 and substituting the parameters with parameters reported in Table S5.3, we can estimate that, for example, externalizing problem scores increased by approximately 0.65 units between age 4 and age 5 (from an expected average score of 6.69 to an expected average score of 7.34) and increased by 0.77 units from age 5 to age 6 (from an expected average score of 7.34 to an expected average score of 8.11), but increased only by 0.11 units between age 8 and 9 years (from an expected average score of 8.99 to an expected average score of 9.10) and started to decrease between age 9 and 10 years, with a decrease of approximately 0.10 units (from an expected average score of 9.10 to an expected average score of 9.00).

**Table S5.3**

*Parameters of the Final Spline Model of Externalizing Problems*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effects | Coefficient | *95% CI* LB | *95% CI* UB | z |
| Intercept | 6.31 | 5.57 | 7.05 | 16.73\*\* |
| Linear trend S1 | 0.38 | -0.24 | 1.00 | 1.19 |
| Linear trend S2 | 0.99 | 0.47 | 1.51 | 3.71\*\* |
| Quadratic trend S2 | -0.11 | -0.19 | -0.03 | -2.62\* |
| Random Effects | Coefficient | *95% CI* LB | *95% CI* UB |  |
| Intercept Variance | 5.20 | 3.13 | 8.63 |  |
| Slope S1 Variance | 2.26 | 1.18 | 4.34 |  |
| Slope S2 Variance | 0.11 | 0.03 | 0.51 |  |
| Within-Individuals Variance | 4.83 | 3.78 | 6.18 |  |

Note. S1 : Spline 1; S2: Spline 2; 95% CI: 95% Confidence Intervals; LB: Lower Bound; UB: Upper Bound. \* *p* < .01; \*\* *p* < .001.

We then added covariates sex at birth (dummy coded as male), age at time of placement (centred at age 2 years) and the standardised SES component score to the model described in Table S5.3. The results of this further model run are reported in Table S5.4: although the had a good fit to the data, Wald *χ2 (7)* = 32.90, *p* < .0001, the covariates did not display a significant association with the outcomes, see Table S5.4. The results indicated that the spline effects did not change substantially once we controlled for other covariates.

**Table S5.4.**

*Parameters of the Final Spline Model of Externalizing Problems with Covariates*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Fixed Effects | Coefficient | *95% CI* LB | *95% CI* UB | z |
| Intercept | 6.06 | 4.72 | 7.41 | 8.83\*\*\* |
| Linear trend S1 | 0.38 | -0.25 | 1.02 | 1.18 |
| Linear trend S2 | 0.98 | 0.46 | 1.51 | 3.68\*\*\* |
| Quadratic trend S2 | -0.11 | -0.19 | -0.03 | -2.60\*\* |
| Sex at Birth: Male | 0.21 | -1.18 | 1.61 | 0.30 |
| Age at Placement: Birth | *Reference* |  |  |  |
| Age at Placement: 1 to <3 yrs | 0.19 | -1.28 | 1.67 | 0.26 |
| Age at Placement: 3 to <5 yrs | 0.11 | -2.34 | 2.56 | 0.09 |
| Std SES Component | 0.30 | -0.44 | 1.03 | 0.79 |
| Random Effects | Coefficient | *95% CI* LB | *95% CI* UB |  |
| Intercept Variance | 5.15 | 3.10 | 8.54 |  |
| Slope S1 Variance | 2.24 | 1.16 | 4.30 |  |
| Slope S2 Variance | 0.12 | 0.03 | 0.52 |  |
| Within-Individuals Variance | 4.82 | 3.77 | 6.17 |  |

Note. S1: Spline 1; S2: Spline 2; 95% CI: 95% Confidence Intervals; LB= Lower Bound; UB= Upper Bound. \* *p* < .05 ; \*\* *p* < .01; \*\*\* *p < .001*.