Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genetic and environmental contributions to curve of Spee morphology: a twin study

Greta Gailiūtė
Lithuanian University of Health Sciences
Arūnas Vasiliauskas
Lithuanian University of Health Sciences
Giedrė Trakinienė
Lithuanian University of Health Sciences
Antanas Šidlauskas
antanas.sidlauskas@lsmu.lt

Lithuanian University of Health Sciences

Article

Keywords: curve of Spee, twins, heritability, occlusion, dental arch, structural equation modelling

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10928873/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Genetic and environmental contributions to curve of Spee morphology: a twin study

Greta Gailiūtė[1], Arūnas Vasiliauskas[1], Giedrė Trakinienė[1] and Antanas Šidlauskas[1]*

[1]Department of Orthodontics, Faculty of Odontology, Medical Academy, Lithuanian University of Health Sciences, J. Lukšos-Daumanto g. 6, LT-50106 Kaunas, Lithuania

Correspondence and requests for materials should be addressed to Antanas Šidlauskas (email: antanas.sidlauskas@lsmu.lt).

Abstract

The curve of Spee is clinically relevant to orthodontic diagnosis and treatment, but the genetic and environmental contributions to its morphology remain unclear. This retrospective cross-sectional classical twin study evaluated curve of Spee depth and sagittal projected area in standardised digital mandibular models from 154 same-sex twin pairs (89 monozygotic and 65 dizygotic) with DNA-confirmed zygosity. Age- and sex-adjusted ACE, AE, CE and E-only models were compared using 1,000-replicate parametric-bootstrap likelihood-ratio tests and Akaike information criterion. Raw-scale AE models were selected for all six outcomes. Within these models, additive genetic variance accounted for 35.1–47.2% of projected-area variance and 27.3–41.1% of depth variance; unique environmental variance, including measurement error, accounted for the remainder. AE-versus-E-only comparisons remained significant for every outcome after Holm correction (adjusted $p = 0.006$). However, the distinction between additive genetic and shared environmental covariance was limited, and natural-log sensitivity analyses favoured CE for right-side area and depth. The findings support familial contributions to curve of Spee morphology and are compatible with moderate additive genetic effects under the raw-scale AE specification, while emphasising model and scale dependence.

Keywords: curve of Spee; twins; heritability; occlusion; dental arch; structural equation modelling

Introduction

The curve of Spee is the anteroposterior curvature of the mandibular occlusal plane and is routinely considered in orthodontic diagnosis and treatment planning. Its depth varies among individuals, and associations with dentoskeletal patterns have been reported, although the certainty of evidence is very low [1]. Long-term stability after orthodontic levelling has been investigated [2]. Nevertheless, the biological basis of interindividual variation in curve of Spee morphology remains incompletely understood.

Dental and occlusal phenotypes reflect interacting genetic and environmental influences, whose relative contributions may vary across traits and stages of development. Systematic reviews have reported substantial heterogeneity in heritability estimates and study methods, with low certainty of evidence for several occlusal outcomes [3, 4]. Recent twin studies have identified genetic contributions – predominantly additive for many, but not all, traits – to permanent dental-arch form [5], mandibular cephalometric dimensions [6], palatal dimensions [7], and longitudinal arch traits [8]. However, these findings cannot be extrapolated directly to curve of Spee morphology.

Direct evidence for familial contributions to the curves of occlusion remains limited. A study of siblings seeking orthodontic treatment reported a high heritability estimate for curve of Spee depth [9]. However, sibling resemblance alone cannot separate additive genetic effects from environmental influences shared within families. In contrast, the classical monozygotic–dizygotic twin design uses differences in expected genetic relatedness and cross-twin covariance between the two zygosity groups to estimate additive genetic, shared environmental and unique environmental variance components [10]. Moreover, the studies cited above characterised curve of Spee morphology primarily using linear depth measurements [1, 9]. Sagittal projected area may provide complementary information by capturing the extent of the curve along the dental arch rather than only its deepest point (Fig. 1). An earlier longitudinal twin thesis examined curve-of-Spee depth and polynomial shape descriptors using broad-sense heritability estimates, but did not fit variance-component structural equation models or assess sagittal projected area [11].

To our knowledge, no peer-reviewed classical twin study has used structural equation modelling to estimate genetic and environmental contributions to both curve of Spee depth and sagittal projected area in twins with DNA-confirmed zygosity. Therefore, this study aimed to estimate the additive genetic, shared environmental and unique environmental variance components for left-side, right-side and bilateral mean curve of Spee depth and sagittal projected area, with adjustment for age and sex.

Results

Participant characteristics

Of 170 potentially eligible twin pairs, 154 same-sex pairs with complete data (308 individuals) were included in the analysis: 89 monozygotic and 65 dizygotic pairs. Sixteen pairs were excluded because one or both members did not meet the study-specific eligibility criteria. The included sample comprised 63 male and 91 female twin pairs. Mean age was 19.34 years (SD, 4.38; range, 13.08–37.83 years) overall, 19.72 years (SD, 4.64) among monozygotic twins and 18.81 years (SD, 3.96) among dizygotic twins (Table 1). Complete data for all six curve of Spee outcomes were available for every included pair.

Measurement reliability

Intra-examiner reliability was excellent for all outcomes. Intraclass correlation coefficients ranged from 0.978 to 0.992 for curve of Spee depth and from 0.994 to 0.998 for sagittal projected area. Dahlberg errors ranged from 0.066 to 0.119 mm for depth and from 0.865 to 1.414 mm² for projected area. Two-sided paired *t*-tests showed no evidence of systematic differences between the initial and repeated measurements (depth, $p = 0.153–0.431$; projected area, $p = 0.630–0.842$; Table 2 and Supplementary Table S1).

Curve of Spee measurements and intrapair correlations

Mean sagittal projected area was 49.18 mm² (SD, 22.23) on the left and 44.76 mm² (SD, 19.34) on the right, with a bilateral mean of 46.97 mm² (SD, 18.11). Mean curve of Spee depth was 2.23 mm (SD, 0.86) on the left and 2.10 mm (SD, 0.79) on the right, with a bilateral mean of 2.17 mm (SD, 0.71; Table 3).

Intrapair correlations were numerically higher in monozygotic than in dizygotic pairs for all outcomes. Correlations ranged from 0.275 to 0.435 in monozygotic pairs and from 0.152 to 0.288 in dizygotic pairs (Table 3). The highest correlation in both groups was observed for bilateral mean sagittal projected area (rMZ = 0.435; rDZ = 0.288). These descriptive correlations were not used for formal inference about differences between the zygosity groups.

Twin-model comparison

All ACE, AE, CE and E-only models converged with OpenMx status code 0. In 1,000-replicate parametric-bootstrap comparisons, removing the shared environmental component did not worsen model fit for any outcome (AE versus ACE: observed likelihood-ratio test statistic = 0.00–0.11; bootstrap p = 0.378–1.000). Removing the additive genetic component yielded bootstrap p values of 0.035–0.254 (CE versus ACE). Although nominal p values were below 0.05 for left-side projected area, left-side depth and bilateral mean depth, none remained significant after Holm correction across the six outcomes (adjusted p = 0.210–0.254). AE had the lowest Akaike information criterion for all six raw-scale outcomes, with AIC values 0.24–3.03 points lower than those of the corresponding CE models. The compatibility-and-AIC selection procedure described in Methods therefore selected AE for every outcome, although the model-comparison pattern indicated limited ability to distinguish additive genetic from shared environmental covariance (Table 4 and Supplementary Table S2).

The E-only models, obtained by constraining additive genetic variance to zero, fitted worse than the corresponding raw-scale AE models for all six outcomes (1,000-replicate bootstrap p = 0.001–0.003). All six comparisons remained significant after Holm correction (adjusted p = 0.006) and Benjamini–Hochberg false-discovery-rate correction (adjusted p = 0.0015–0.0030). Thus, evidence against the E-only model was present under the AE specification, whereas the distinction between additive genetic and shared environmental covariance remained unresolved.

Standardised variance components

Within the selected raw-scale AE models, standardised additive genetic variance estimates ranged from 35.1% (95% profile-likelihood CI, 16.9–50.7%) to 47.2% (95% profile-likelihood CI, 30.2–60.9%) for sagittal projected-area outcomes and from 27.3% (95% profile-likelihood CI, 8.4–44.0%) to 41.1% (95% profile-likelihood CI, 23.0–56.2%) for depth outcomes. The complementary unique environmental component, which also includes measurement error, accounted for 52.8–64.9% of projected-area variance and 58.9–72.7% of depth variance (Table 4). The confidence intervals were obtained using direct profile-likelihood analyses, and their endpoints reproduced the 95% likelihood-ratio target to within 0.0001. Standardised additive genetic variance estimates and their profile-likelihood confidence intervals are presented in Fig. 2 and Supplementary Table S3.

Scale-sensitivity analysis

All ACE, AE, CE and E-only models also converged after natural-log transformation of the six positive outcomes. AE retained the lowest AIC for left-side projected area, bilateral mean projected area, left-side depth and bilateral mean depth. CE had the lowest AIC for right-side projected area (440.81 versus 441.88 for AE) and right-side depth (319.14 versus 320.44 for AE). Standardised additive genetic variance estimates from the log-scale AE models ranged from 0.243 to 0.376. Thus, the lowest-AIC model for the two right-side outcomes was sensitive to outcome scale. The raw-scale analysis was retained as the primary analysis because it preserves the original measurement units, whereas the natural-log analysis was treated as a sensitivity analysis (Supplementary Table S4).

Study-size sensitivity analysis

In the scenario-based study-size sensitivity analysis, the fixed sample of 154 complete twin pairs yielded estimated power values of 67.7%, 82.9%, 93.8%, 98.1% and 99.6% for hypothetical additive genetic components of A = 0.20, 0.25, 0.30, 0.35 and 0.40, respectively (Supplementary Table S5). An additional sample-size grid for A = 0.30 is presented in Supplementary Table S6. These estimates describe the probability of detecting A under the assumed AE data-generating model. They should not be interpreted as retrospective power estimates for the observed effects or as evidence of comparable ability to distinguish A from C.

Discussion

In this classical twin study with DNA-confirmed zygosity, monozygotic intrapair correlations were numerically higher than dizygotic correlations for all six curve of Spee phenotypes. Raw-scale AE models had the lowest AIC for every outcome and were selected using the bootstrap-compatibility-and-AIC procedure described in Methods. Within these models, standardised additive genetic variance estimates ranged from 27.3% to 47.2%, whereas unique environmental variance, including measurement error, accounted for 52.8–72.7%. AE-versus-E-only comparisons remained significant after Holm correction. However, model comparisons did not distinguish additive genetic from shared environmental covariance after correction for multiple testing across the six outcomes, AIC differences between AE and CE were modest, and log-scale sensitivity analyses favoured CE for right-side projected area and depth. The findings therefore support familial covariance and are compatible with moderate additive genetic contributions under the raw-scale AE specification, but they do not unequivocally distinguish additive genetic from shared environmental covariance.

These findings extend the limited evidence concerning familial influences on the curves of occlusion. Al-Qawasmi and Coe reported a high heritability estimate for curve of Spee depth among siblings seeking orthodontic treatment [9]. The present estimates should not be interpreted as a direct replication of that sibling-based finding because sibling resemblance may reflect both genetic effects and environments shared within families, whereas the classical twin design estimates variance components from differences in covariance between monozygotic and dizygotic twins [10]. Differences in participant age, source population, imaging modality, phenotype definition and analytical approach may also contribute to variation among reported estimates [3, 4]. Recent geometric-morphometric twin studies provide informative but non-equivalent comparators. Lin et al. analysed 64 monozygotic and 38 dizygotic twins using three-dimensional landmark-based dental models. AE models were the most parsimonious for most univariate shape components, with a CE model selected for maxillary PC2; the multivariate model supported additive genetic and unique environmental influences, and heritability estimates ranged from 0.61 to 0.74 [12]. Longitudinal polynomial modelling across primary, mixed and permanent dentitions likewise showed substantial but developmentally fluctuating genetic influences on dental-arch shape [13], while a 2026 longitudinal study of 188 twin pairs explicitly included steepness of the curve of Spee within dental-arch shape components and predominantly selected AE models, with a CE exception for primary-dentition mandibular PC3; heritability estimates for AE-modelled components were 0.60–0.84 [14]. Because these outcomes represent multivariate whole-arch

shape rather than isolated curve depth or sagittal projected area, their higher estimates are not direct replications of the present findings.

The standardised additive genetic variance estimates from the selected AE models were generally lower than the moderate-to-high genetic contributions reported for several permanent dental-arch, palatal and mandibular cephalometric traits [5–8]. Genetic contributions have also been reported for third-molar root mineralisation in twins [15], although that developmental phenotype is not directly comparable with curve of Spee morphology. Heritability estimates are specific to the phenotype, population, developmental stage, environmental context and fitted model [3, 4, 8, 16]. Age- and sex-related evidence should, however, be distinguished from moderation of variance components. Hughes et al. included both same-sex and opposite-sex dizygotic twin pairs and found sexual heterogeneity in dental-arch shape, particularly in the permanent dentition [13]. In a 2024 cross-sectional digital-scan study of adults aged 18–65 years, curve of Spee depth was positively associated with overjet but did not differ significantly between females and males [17]. The longitudinal Iowa Facial Growth Study showed eruption-related increases in curve depth from childhood through second-molar eruption, followed by relative stability from adolescence into early adulthood, without significant differences between sides or sexes [18]. Although age and sex were included as covariates in the present analysis, this adjustment accounts for mean differences but does not test whether the variance components differ according to age or sex [16]. Curve of Spee depth has been associated with overjet [17] and eruption stage [18]; the roles of occlusal contacts and functional loading remain hypotheses. These potential mechanisms were not measured in this study and therefore were not evaluated as mediators. The estimated variance components are population-level and model-dependent quantities, not evidence for specific genes, exposures or biological immutability.

Sagittal projected area complemented the conventional depth measurement by incorporating the extent of the curve along the standardised sagittal projection. The highest additive genetic variance estimate was observed for bilateral mean projected area and the lowest for right-side depth. However, the confidence intervals overlapped, and no equality-constrained or multivariate model was fitted to test differences among outcomes or between sides. Because the left-side, right-side and bilateral mean measures were correlated, these estimates do not demonstrate that projected area is more heritable than depth or that one side is under stronger genetic control. The value of projected area lies in its broader phenotypic characterisation. Its construct validity, inter-examiner reproducibility and clinical interpretability should be evaluated in independent samples. Recent work has shown that automated landmark detection can measure curve of Spee depth from intraoral scans with a small median

measurement error (0.05 mm), suggesting a route to more standardised and scalable assessment [19]. However, that method assessed depth rather than projected area and therefore does not establish the construct validity of the present area phenotype.

Within the selected AE models, the estimated E component exceeded A for every phenotype. In classical twin analysis, E captures non-shared environmental influences and measurement error [4]. The excellent repeat-measurement reliability and small Dahlberg errors suggest that measurement imprecision alone is unlikely to explain the E estimates, although they do not exclude errors arising from model orientation, image projection or examiner-specific landmark selection. The E component also does not identify specific environmental determinants. A 2024 cross-sectional study of 231 adults reported differences in bite force and chewing ability across curve of Spee categories, with the highest mean bite force in participants with a deep curve; because the design was cross-sectional, directionality and mediation cannot be established [20]. Functional asymmetry, including preferred chewing side, is one possible source of within-pair variation. Twin studies have examined preferred chewing side in relation to other lateralities [21] and to facial asymmetry [22]. However, chewing laterality was not measured in the present study, and any relationship with curve of Spee morphology remains hypothetical.

The clinical implications should be interpreted cautiously. The findings support curve of Spee morphology as a multifactorial characteristic. However, specific environmental exposures, orthodontic treatment response and post-treatment stability were not assessed; therefore, the observed E component cannot be used to identify modifiable factors or predict individual treatment response. Longitudinal studies incorporating eruption, treatment, occlusal and functional data are required before conclusions about modifiability, prognosis or individualised care can be drawn.

Strengths. This study included 154 same-sex twin pairs with complete outcome data, DNA-based zygosity determination, conventional depth and sagittal projected-area phenotypes, excellent intra-examiner repeatability and adjustment for age and sex. The statistical workflow included explicit comparison of ACE, AE, CE and E-only models, 1,000-replicate parametric-bootstrap likelihood-ratio tests, Holm correction across the six AE-versus-E-only comparisons, direct profile-likelihood confidence intervals and a natural-log scale-sensitivity analysis. All observed-data models converged, and all 18,000 requested bootstrap replicates were successful. The scenario-based study-size sensitivity analysis indicated at least 80% simulated power for $A \geq 0.25$ under the assumed AE model. However, because power in classical twin designs depends on the assumed variance structure and the specific

model comparison [23], this analysis did not establish comparable ability to distinguish A from C in the full ACE model.

Limitations. The retrospective cross-sectional design precludes inference about development, temporal change or causation. The age range (13.08–37.83 years) spans late dental maturation and adulthood. Although longitudinal evidence suggests relative stability of curve depth after adolescence [18], the present cross-sectional analysis did not model nonlinear age effects, developmental stages or age-specific variance components; sex-specific variance components were likewise not estimated. Sagittal projected area is a two-dimensional representation derived from a three-dimensional model and is not yet a standardised curve of Spee measure. Inter-examiner reproducibility was not assessed, and the effects of model orientation, image projection and manual delineation require further evaluation. Candidate non-shared exposures, including chewing preference and occlusal function, were unavailable; consequently, E cannot be attributed to specific mechanisms. In addition, a reason-specific screening log was unavailable for the 16 excluded twin pairs. Moreover, several of the closest comparator studies were drawn from Australian/Adelaide twin cohorts that may overlap; independence across samples cannot be established from the published reports [12–14]. The examiner was not blinded to zygosity because it was visible in the source worksheet, although zygosity was not consulted during model orientation or outcome measurement.

No single primary outcome was designated, and six correlated outcomes were analysed. Holm correction was applied across the six outcomes within the specified model-comparison families, but this does not eliminate the interpretative complexity arising from correlated outcomes and multiple related model-selection comparisons. Accordingly, interpretation should emphasise effect estimates, confidence intervals and sensitivity analyses rather than statistical significance alone. Natural-log transformation changed the lowest-AIC model from AE to CE for right-side projected area and depth, demonstrating scale dependence. The classical twin model also relies on the equal-environments and random-mating assumptions, and gene–environment correlation and interaction were not modelled. Finally, the sample had limited ability to distinguish A from C, particularly for right-side depth, and recruitment from a single institutional twin-centre and clinical population limits generalisability.

Conclusions

Curve of Spee depth and sagittal projected area showed numerically greater intrapair similarity among monozygotic than among dizygotic twins in this sample with DNA-confirmed zygosity. Raw-scale AE models had the lowest AIC for all six outcomes and yielded standardised additive genetic variance estimates of 27.3–47.2%; AE-versus-E-only comparisons remained significant after Holm correction. However, the ability to distinguish additive genetic from shared environmental covariance was limited, and the lowest-AIC model for right-side projected area and depth was sensitive to logarithmic transformation. The findings therefore support a multifactorial and partly familial basis for curve of Spee morphology and are compatible with moderate additive genetic contributions under the primary raw-scale AE specification. Further longitudinal studies and independent replication are required to validate sagittal projected area and identify factors associated with within-pair variation.

Methods

Study design and setting

This retrospective cross-sectional classical twin study evaluated genetic and environmental contributions to variation in curve of Spee morphology. Participants were identified from the Twin Centre database of the Lithuanian University of Health Sciences. The study used three-dimensional digital mandibular dental models obtained during initial orthodontic consultations at the Department of Orthodontics, Lithuanian University of Health Sciences. The dental models were collected and digitised between 2011 and 2019. Data selection, model review, image preparation and primary outcome measurements were performed in January 2026. Before analysis, all records were assigned study-specific codes. The same study sample and dataset of curve of Spee measurements formed the basis of the first author's Lithuanian-language Master's thesis [24]. The present manuscript reports an expanded reanalysis, including the additional model comparisons, bootstrap inference, multiplicity adjustments and sensitivity analyses described below.

Participants

The source database contained 170 potentially eligible twin pairs. After application of the study-specific eligibility criteria, 154 same-sex twin pairs with complete and analysable data for both members (308 individuals) were included: 89 monozygotic pairs (57.8%) and 65 dizygotic pairs (42.2%). Sixteen pairs were excluded because one or both members did not meet the eligibility criteria; a reason-specific screening log was not available for the present analysis. The included sample comprised 63 male and 91 female twin pairs. Mean participant age was 19.34 years (SD, 4.38 years).

Eligibility required confirmed zygosity, high-quality mandibular digital dental models, eruption of all permanent teeth except third molars, complete data for both members of the twin pair and no history of orthodontic treatment. Exclusion criteria were previous orthodontic treatment; missing, unerupted, extracted or traumatically lost permanent teeth other than third molars; prosthetic restorations or extensive coronal restorations affecting the measurement region; major skeletal anomalies, syndromes or facial trauma; and digital models with deformation, scanning artefacts or insufficient anatomical detail.

Zygosity determination

Zygosity information was obtained from the Lithuanian University of Health Sciences Twin Centre database. Zygosity was determined at the certified laboratory UAB Synlab Lithuania. DNA was extracted from blood samples using the Genomic DNA Extraction Module and Chelex-100 chelating resin. Fifteen polymorphic short tandem repeat loci – D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA – together with an amelogenin fragment were amplified by polymerase chain reaction using the AmpFℓSTR® Identifiler® system (Applied Biosystems, USA). Zygosity was assigned based on concordance across the tested loci.

Digital-model preparation

Dental impressions were obtained using Kromopan alginate material (LASCOD, Italy) and poured in type IV dental stone (Marmorock N, Siladent). The resulting casts were digitised using a 3Shape E3 laboratory scanner and 3Shape Ortho System software (3Shape A/S, Copenhagen, Denmark). De-identified mandibular models were available for the present analysis in stereolithography (STL) format. Each model was inspected according to predefined quality criteria and retained only when all required anatomical landmarks were visible and no scanning artefacts, deformation or data loss that could affect measurement was present.

Each mandibular model was manually oriented in the sagittal plane by the same examiner using 3Shape 3D Viewer. Separate sagittal screenshots were generated for the right and left sides of each model and analysed in ImageJ version 1.54p (National Institutes of Health, Bethesda, MD, USA). Each screenshot was calibrated individually by entering the distance displayed between the two blue calibration markers in 3Shape 3D Viewer as the known reference distance using the Set Scale function in ImageJ. Linear measurements were expressed in millimetres and projected areas in square millimetres. The same manual orientation and measurement sequence was applied to all models. The examiner was not blinded to zygosity because it was visible in the source worksheet; however, zygosity was not consulted during model orientation or outcome measurement.

Curve of Spee outcomes

Curve of Spee depth and sagittal projected area were measured separately on the right and left sides of the mandibular arch (Fig. 1). For each side, a reference line connected the cusp tip of the mandibular

canine to the distobuccal cusp tip of the mandibular second molar. The curve contour was manually traced from the mandibular canine cusp tip through the visible buccal cusp tips of the premolars and molars to the distobuccal cusp tip of the mandibular second molar. Depth was defined as the greatest perpendicular distance from the deepest point of the occlusal curve to the reference line. Sagittal projected area was defined as the two-dimensional area bounded by the curve of Spee contour and the reference line in the standardised sagittal projection. The area was manually delineated using the polygon-selection tool in ImageJ. Bilateral mean depth and bilateral mean projected area were calculated as the arithmetic means of the corresponding right- and left-side measurements. For reproducibility, the ImageJ workflow used the Straight Line tool for the reference line, Polygon Selection for manual area delineation, and Measure for projected-area calculation.

The term sagittal projected area is used deliberately because this measure represents a two-dimensional projection rather than the complete three-dimensional geometry of the occlusal curve. No single primary outcome was designated. All six left-side, right-side and bilateral mean depth and projected-area variables were analysed as study outcomes; none was selected on the basis of its observed effect estimate or statistical significance.

Measurement reliability

Twenty individual mandibular models were selected at random for repeat assessment by the same examiner one month later. The complete orientation and measurement workflow was repeated: each model was reoriented in 3Shape 3D Viewer using the same protocol as in the primary assessment, new right- and left-side sagittal screenshots were generated and individually calibrated, and all depth and projected-area outcomes were remeasured in ImageJ. The examiner did not have access to the initial measurements during reassessment. Intra-examiner reliability was evaluated using a two-way mixed-effects, absolute-agreement, single-measure intraclass correlation coefficient [25]. Random error was quantified using Dahlberg's formula, and systematic differences between the initial and repeated measurements were assessed using a two-sided paired Student's *t*-test. Statistical significance was defined as $p < 0.05$.

Statistical analysis

Descriptive analyses were performed using IBM SPSS Statistics version 31.0.2.0 and independently reproduced in R version 4.5.2. Genetic structural equation models, profile-likelihood calculations,

parametric-bootstrap tests and scenario-based simulations were performed in R using OpenMx version 2.22.10; data were imported using readxl version 1.4.5. Continuous variables were summarised using the mean, standard deviation, median, interquartile range, minimum and maximum, and categorical variables using counts and percentages. Age was converted from months to years and centred at the sample mean. Sex was coded as 0 for male and 1 for female. Because co-twins had the same age and sex, both covariates were entered into the expected-mean structure for each member of the pair.

Outcome distributions and potential outliers were assessed using histograms, quantile–quantile plots, the Shapiro–Wilk test and the 1.5 × interquartile-range criterion. No additional observations were excluded from the analysis on the basis of these distributional or outlier diagnostics. Although formal tests indicated deviations from normality, graphical inspection suggested only mild-to-moderate asymmetry. The primary models were fitted to the original measurement scale, and all models were repeated after natural-log transformation as a scale-sensitivity analysis. Pearson intrapair correlations were calculated separately for monozygotic and dizygotic pairs. Falconer's formula, $h^2 = 2(r\text{MZ} - r\text{DZ})$, was used only as a descriptive preliminary estimate; principal inference was based on structural equation modelling.

For each outcome, a univariate classical twin model partitioned phenotypic variance into additive genetic (A), shared environmental (C) and unique environmental (E) components, with E also including measurement error. Expected cross-twin covariance was specified as A + C for monozygotic pairs and 0.5A + C for dizygotic pairs. The models assumed random mating, equivalent effects of shared environmental factors on monozygotic and dizygotic twin similarity (the equal-environments assumption), and no modelled gene–environment correlation or interaction [10, 16]. Age and sex were included as covariates in the mean structure.

For each outcome, the full ACE model was compared with the nested AE and CE models, and the AE model was compared with the E-only model. Because variance components fixed at zero lie on the boundary of the parameter space [26], final model-comparison p values were obtained using parametric-bootstrap likelihood-ratio tests rather than conventional asymptotic chi-square reference distributions. For each comparison, 1,000 datasets were generated under the fitted null model, both the null and alternative models were refitted, and a plus-one-corrected bootstrap p value was calculated. A reduced model was considered compatible with ACE when its unadjusted bootstrap p value was at least 0.05. Among the full ACE model and the compatible reduced AE and CE models, the model with the lowest Akaike information criterion was selected. Conventional asymptotic likelihood-ratio p values were retained only as diagnostic outputs.

Standardised A and E estimates were reported from the selected raw-scale AE models. Final 95% confidence intervals were obtained by direct profile-likelihood optimisation using the 3.841 likelihood-ratio criterion. The six AE-versus-E-only comparisons constituted the principal cross-outcome inferential family and were adjusted using Holm's method; Benjamini–Hochberg false-discovery-rate adjustment was reported as a sensitivity analysis. Holm and Benjamini–Hochberg adjustments were also examined within the six AE-versus-ACE and six CE-versus-ACE comparison families, whereas their unadjusted bootstrap *p* values were used only for the model-selection procedure described above. All tests used $\alpha = 0.05$. Reproducibility criteria required OpenMx status code 0 for all observed-data models, finite direct-profile endpoints reproducing the target likelihood ratio within 0.0001, and at least 90% successful bootstrap refits. All 154 included twin pairs had complete analysis data, and no imputation was performed.

Study-size sensitivity analysis

The study sample comprised all twin pairs available in the source database who met the eligibility criteria and had complete analysable data for both members; the sample size was therefore fixed. A scenario-based Monte Carlo sensitivity analysis characterised the power of the fixed sample under a set of hypothetical AE data-generating models [23]. Standardised models specified A values from 0.20 to 0.40 and $E = 1 - A$, retained the observed monozygotic-to-dizygotic allocation ratio and generated 5,000 datasets per scenario. In each simulated dataset, additive genetic variance was tested by comparing the nested AE and E-only models. Because the null hypothesis fixes the single non-negative variance component at the boundary of the parameter space ($A = 0$), significance was determined using the asymptotic 50:50 mixture of χ^2_0 and χ^2_1 distributions, corresponding to a likelihood-ratio critical value of 2.7055 at $\alpha = 0.05$ [26]. This analysis assessed the detection of A under the assumed AE model and did not establish equivalent ability to distinguish A from C in the full ACE model.

Ethics approval and consent to participate

The study protocol was approved by the Kaunas Regional Biomedical Research Ethics Committee (No. BE-2-41) and Bioethics Centre of the Lithuanian University of Health Sciences (No. 2026-BEC2-0282). Written informed consent was obtained from all participants and, for participants younger than 18 years, from a parent or legal guardian. The study was conducted in accordance with the Declaration of Helsinki and relevant national and institutional guidelines and regulations. Only study-coded data without direct personal identifiers were used for the present analysis.

Use of artificial intelligence-assisted tools

During the preparation of this manuscript, ChatGPT (OpenAI) was used to assist with translation from Lithuanian into English and with language editing, including improvements to sentence structure, clarity and readability. The authors critically reviewed and revised all AI-assisted text and independently verified the accuracy of the scientific content, numerical results, statistical interpretations and references. AI did not determine the study conclusions. The authors take full responsibility for the final manuscript.

Data availability

The study-coded participant-level dataset for the 154 twin pairs supporting the genetic analyses is not publicly available because participant consent does not permit unrestricted sharing and data-protection requirements apply. These data may be made available on reasonable request to the corresponding author, Antanas Šidlauskas (antanas.sidlauskas@lsmu.lt), subject to institutional approval and a data-use agreement. Requests will receive an initial response within 30 days. The study-coded repeat-measurement dataset for the 20 individual mandibular models used in the intra-examiner reliability analysis is included in Supplementary Data 1. The same data are also publicly available in Appendix 1 (p. 41) of Greta Gailiūtė's Master's thesis through the institutional repository of the Lithuanian University of Health Sciences: https://lsmu.lt/cris/entities/etd/fbc2f522-12ce-4ce5-a925-aa63053e858e.

Code availability

The complete R analysis script used for data validation, measurement-reliability calculations, genetic structural equation modelling, profile-likelihood calculations, parametric-bootstrap tests, multiplicity adjustments and scenario-based simulations is provided as Supplementary Software 1. Key analysis outputs, including model diagnostics, bootstrap summaries, multiplicity-adjusted inference, sensitivity analyses and session information, are provided as Supplementary Data 1. The participant-level twin-pair dataset required for the genetic analyses is governed by the controlled-access conditions stated under Data availability and is not included in the public supplementary files. The study-coded repeat-measurement dataset required for the reliability analysis is included in Supplementary Data 1. The same data are also publicly available in Appendix 1 (p. 41) of Greta Gailiūtė's Master's thesis through

the institutional repository of the Lithuanian University of Health Sciences: https://lsmu.lt/cris/entities/etd/fbc2f522-12ce-4ce5-a925-aa63053e858e.

References

[1] Paes-Souza, S. de A. et al. Is there variation in the depth of the curve of Spee in individuals with different dentoskeletal patterns? A systematic review with meta-analysis. *Eur. J. Orthod.* **44,** 491–502; 10.1093/ejo/cjab078 (2022).

[2] Busenhart, D. M., Schätzle, M., Eliades, T. & Papageorgiou, S. N. Long-term stability of curve of Spee depth among orthodontically treated patients: a retrospective longitudinal study. *Orthod. Craniofac. Res.* **27,** 572–581; 10.1111/ocr.12756 (2024).

[3] Giri, J. et al. Heritability of dental arches and occlusal characteristics: a systematic review and meta-analysis. *Eur. J. Orthod.* **45,** 854–867; 10.1093/ejo/cjad061 (2023).

[4] Santana, L. G., Flores-Mir, C., Iglesias-Linares, A., Pithon, M. M. & Marques, L. S. Influence of heritability on occlusal traits: a systematic review of studies in twins. *Prog. Orthod.* **21,** 29; 10.1186/s40510-020-00330-8 (2020).

[5] Lin, T. H., Hughes, T. & Meade, M. J. The genetic and environmental contributions to variation in the permanent dental arch form: a twin study. *Eur. J. Orthod.* **45,** 868–874; 10.1093/ejo/cjad054 (2023).

[6] Šidlauskas, M. et al. Heritability of mandibular cephalometric variables in twins with completed craniofacial growth. *Eur. J. Orthod.* **38,** 493–502; 10.1093/ejo/cjv062 (2016).

[7] Šidlauskienė, M. et al. Genetic and environmental impact on variation in the palatal dimensions in permanent dentition: a twin study. *Sci. Rep.* **14,** 19785; 10.1038/s41598-024-70985-4 (2024).

[8] Giri, J. et al. Genetic and environmental contributions to the development of dental arch traits: a longitudinal twin study. *Eur. J. Orthod.* **47,** cjaf018; 10.1093/ejo/cjaf018 (2025).

[9] Al-Qawasmi, R. & Coe, C. Genetic influence on the curves of occlusion in children seeking orthodontic treatment. *Int. Orthod.* **19,** 82–87; 10.1016/j.ortho.2020.12.003 (2021).

[10] Rijsdijk, F. V. & Sham, P. C. Analytic approaches to twin data using structural equation models. *Brief. Bioinform.* **3,** 119–133; 10.1093/bib/3.2.119 (2002).

[11] Gagliardi, A. *The Development of the Curve of Spee in Australian Twins.* Doctor of Clinical Dentistry thesis, The University of Adelaide (2013).

[12] Lin, T. H., Meade, M. J. & Hughes, T. Dental arch shape in twins: a morphometric study of genetic influences. *Am. J. Orthod. Dentofacial Orthop.* **166,** 583–594; 10.1016/j.ajodo.2024.07.021 (2024).

[13] Hughes, T., Mahmood, Z., Giri, J., Townsend, G. & Brook, A. H. Fluctuating genetic influences at three different stages of development of dental arches: a complex system. *Genes* **16,** 189; 10.3390/genes16020189 (2025).

[14] Giri, J. et al. Genetic and environmental influences on dental arch shape during development: a longitudinal twin study. *Orthod. Craniofac. Res.*; 10.1111/ocr.70162 (2026).

[15] Trakinienė, G. et al. Genetic and environmental influences on third molar root mineralization. *Arch. Oral Biol.* **98,** 220–225; 10.1016/j.archoralbio.2018.11.026 (2019).

[16] Purcell, S. Variance components models for gene–environment interaction in twin analysis. *Twin Res.* **5,** 554–571; 10.1375/136905202762342026 (2002).

[17] Arcas, V. C. et al. Correlation between depth of the curve of Spee and some specific malocclusion characteristics in a population from Sibiu County, Romania–a cross-sectional study. *J. Clin. Med.* **13,** 2750; 10.3390/jcm13102750 (2024).

[18] Marshall, S. D. et al. Development of the curve of Spee. *Am. J. Orthod. Dentofacial Orthop.* **134,** 344–352; 10.1016/j.ajodo.2006.10.037 (2008).

[19] Ma, S. et al. Clinical advances in curve of Spee assessment: deep learning for automatic tooth landmark detection in Invisalign. *Am. J. Orthod. Dentofacial Orthop.* **169,** 110–125.e4; 10.1016/j.ajodo.2025.09.009 (2026).

[20] Alkhalaf, Z. A. et al. Unveiling the influence of the curve of Spee on bite force and chewing ability: a comparative study. *Int. J. Clin. Pract.* **2024,** 6533841; 10.1155/2024/6533841 (2024).

[21] Pihlaja, K. et al. Relations of laterality and chewing sidedness in twins. *Eur. J. Orthod.* **41,** 244–249; 10.1093/ejo/cjy050 (2019).

[22] Heikkinen, E. V., Vuollo, V., Harila, V., Sidlauskas, A. & Heikkinen, T. Facial asymmetry and chewing sides in twins. *Acta Odontol. Scand.* **80,** 197–202; 10.1080/00016357.2021.1985166 (2022).

[23] Verhulst, B. A power calculator for the classical twin design. *Behav. Genet.* **47,** 255–261; 10.1007/s10519-016-9828-9 (2017).

[24] Gailiūtė, G. Genetikos ir aplinkos veiksnių įtaka Spee kreivės gyliui ir plotui: dvynių tyrimas. Master's thesis, Lithuanian University of Health Sciences, Kaunas, Lithuania (2026). https://lsmu.lt/cris/entities/etd/fbc2f522-12ce-4ce5-a925-aa63053e858e.

[25] Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15,** 155–163; 10.1016/j.jcm.2016.02.012 (2016).

[26] Visscher, P. M. A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Res. Hum. Genet.* **9,** 490–495; 10.1375/183242706778024928 (2006).

Acknowledgements

The authors sincerely thank all twins who participated in this study and contributed to the research.

Author contributions

Greta Gailiūtė: investigation (including outcome measurement), data curation, formal analysis, visualisation and writing-original draft.

gretgail0803@kmu.lt

gretagailiute5@gmail.com

Arūnas Vasiliauskas: material collection, study design and methodology.

arunas.vasiliauskas@lsmu.lt

Giedrė Trakinienė: material collection, study design and methodology.

giedre.trakiniene@lsmu.lt

Antanas Šidlauskas: conceptualisation, methodology, resources, supervision and writing-review and editing.

antanas.sidlauskas@lsmu.lt

All authors reviewed and approved the submitted version and agreed to be accountable for their contributions to the work.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1. Standardised measurement of curve of Spee depth and sagittal projected area. (a) Standardised sagittal orientation of the mandibular model. (b) Reference line connecting the mandibular canine cusp tip and the distobuccal cusp tip of the mandibular second molar. (c) Maximum perpendicular depth from the deepest point of the occlusal curve to the reference line. (d) Sagittal projected area bounded by the curve of Spee contour and the reference line, manually delineated in ImageJ. The blue markers and the displayed distance represent the reference measurement used to calibrate the image scale in ImageJ.

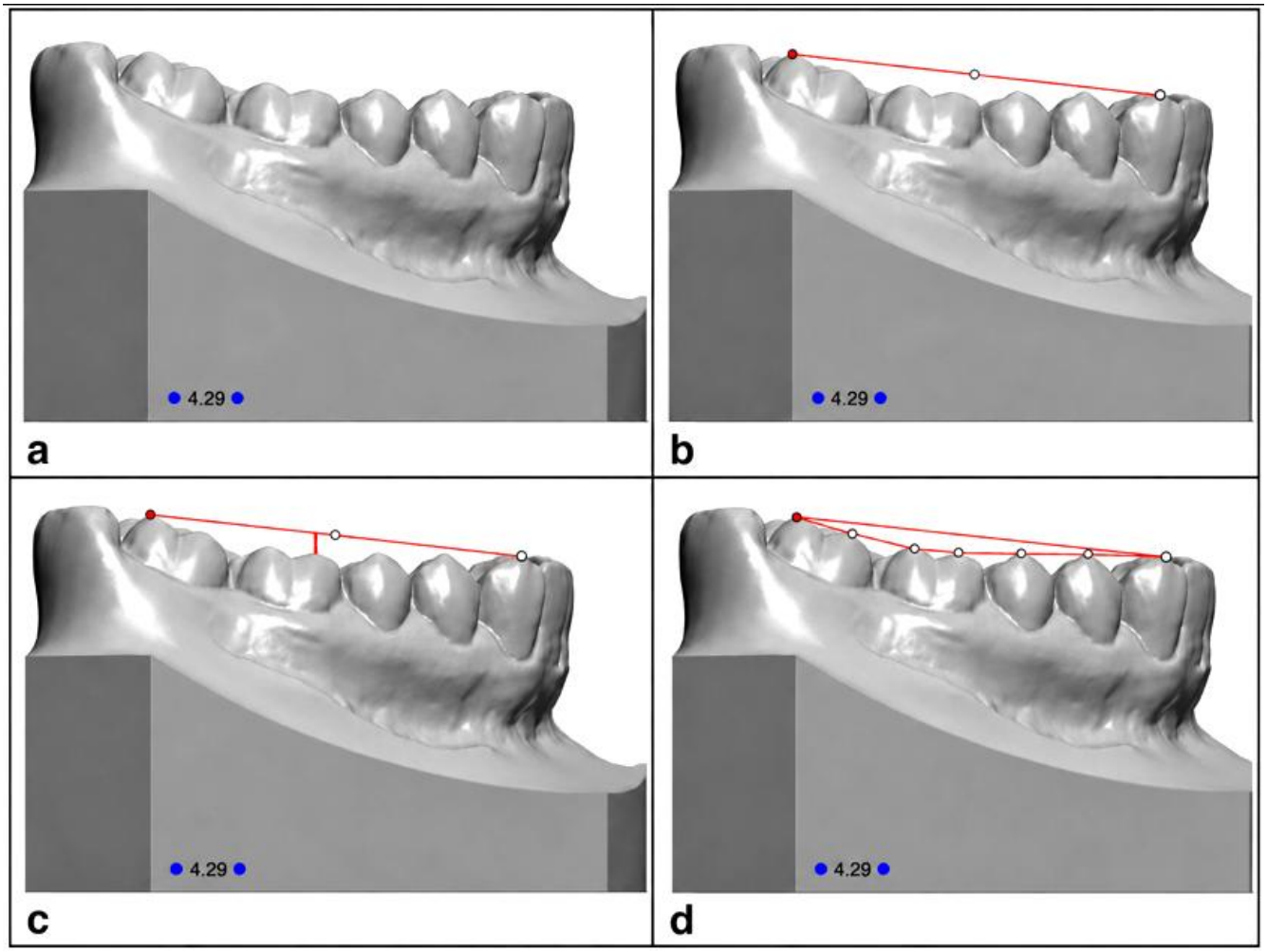
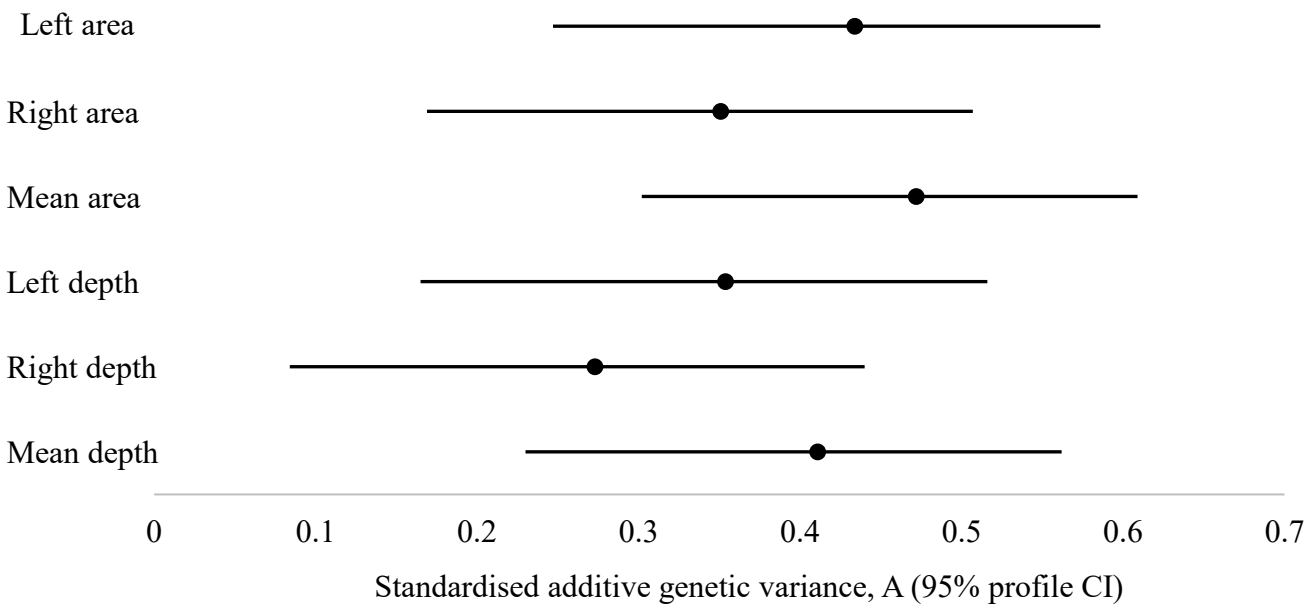


Figure 2. Standardised additive genetic variance estimates for curve of Spee phenotypes. Points represent estimates from the selected age- and sex-adjusted raw-scale AE models; horizontal lines indicate direct profile-likelihood 95% confidence intervals. These estimates describe variance within the study population and should not be interpreted as individual-level genetic determination.



Tables

Table 1. Participant characteristics

Characteristic	Overall	MZ	DZ
Potentially eligible pairs, n	170	–	–
Excluded pairs, n	16	–	–
Twin pairs, n (%)	154 (100)	89 (57.8)	65 (42.2)
Individuals, n	308	178	130
Male pairs, n	63	34	29
Female pairs, n	91	55	36
Age, years, mean (SD)	19.34 (4.38)	19.72 (4.64)	18.81 (3.96)
Age range, years	13.08–37.83	13.08–36.42	13.17–37.83

Abbreviations: DZ, dizygotic; MZ, monozygotic; SD, standard deviation.

Table 2. Intra-examiner measurement reliability

Outcome group	ICC range	Dahlberg error	Systematic error
Depth	0.978–0.992	0.066–0.119 mm	$p = 0.153$–0.431
Sagittal projected area	0.994–0.998	0.865–1.414 mm²	$p = 0.630$–0.842

ICC, two-way mixed-effects, absolute-agreement, single-measure intraclass correlation coefficient. Systematic differences between the initial and repeated measurements were assessed using a two-sided paired *t*-test in 20 remeasured models. Complete outcome-specific ICC estimates, Dahlberg errors and paired-test results are provided in Supplementary Table S1.

Table 3. Descriptive statistics and intrapair correlations

Outcome	Mean (SD)	Median	Range	*r*MZ (*p*)	*r*DZ (*p*)
Area, left (mm²)	49.18 (22.23)	46.64	3.32–140.94	0.378 (0.0003)	0.243 (0.0513)
Area, right (mm²)	44.76 (19.34)	43.04	3.46–123.47	0.358 (0.0006)	0.192 (0.1261)
Area, bilateral mean (mm²)	46.97 (18.11)	44.70	4.75–109.32	0.435 (0.00002)	0.288 (0.0199)
Depth, left (mm)	2.23 (0.86)	2.16	0.29–5.67	0.346 (0.0009)	0.152 (0.2271)
Depth, right (mm)	2.10 (0.79)	2.03	0.38–4.79	0.275 (0.0090)	0.192 (0.1260)
Depth, bilateral mean (mm)	2.17 (0.71)	2.08	0.64–5.05	0.395 (0.0001)	0.200 (0.1094)

Abbreviations: DZ, dizygotic; MZ, monozygotic; SD, standard deviation. Pearson intrapair correlations are based on 89 MZ and 65 DZ pairs. Two-sided, unadjusted *p* values test whether each correlation differs from zero. No formal tests comparing MZ and DZ correlations were performed.

Table 4. Selected raw-scale AE-model fit and standardised variance components

Outcome	AE AIC	Bootstrap *p**	A (95% CI)	E (95% CI)
Area, left	2771.24	0.001	0.434 (0.247–0.586)	0.566 (0.414–0.753)
Area, right	2691.60	0.002	0.351 (0.169–0.507)	0.649 (0.493–0.831)
Area, bilateral mean	2639.20	0.001	0.472 (0.302–0.609)	0.528 (0.391–0.698)
Depth, left	775.41	0.001	0.354 (0.165–0.516)	0.646 (0.484–0.835)
Depth, right	726.90	0.003	0.273 (0.084–0.440)	0.727 (0.560–0.916)
Depth, bilateral mean	655.77	0.001	0.411 (0.230–0.562)	0.589 (0.438–0.770)

Abbreviations: A, additive genetic variance; AIC, Akaike information criterion; CI, confidence interval; E, unique environmental variance, including measurement error. The asterisk identifies plus-one-corrected *p* values from 1,000-replicate parametric-bootstrap comparisons of AE versus E-only; all six comparisons remained significant after Holm correction (adjusted $p = 0.006$). The A and E intervals are direct profile-likelihood 95% confidence intervals. Complete raw-scale ACE, AE, CE and E-only model-fit statistics are provided in Supplementary Table S2; natural-log sensitivity results are provided in Supplementary Table S4.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx