Harmonizing behavioral outcomes across studies, raters, and countries: application to the genetic analysis of aggression in the ACTION Consortium

Justin M. Luningham[1,2], **Anne M. Hendriks**[3,4], **Eva Krapohl**[5], **Hill Fung Ip**[3,4], **Catharina E.M. van Beijsterveldt**[3,4], **Sebastian Lundström**[6,7], **Eero Vuoksimaa**[8], **Tellervo Korhonen**[8], **Paul Lichtenstein**[9], **Robert Plomin**[5], **Lea Pulkkinen**[10], **Richard J. Rose**[11], **Jaakko Kaprio**[8,12], **Meike Bartels**[3,4,13], **Dorret I. Boomsma**[3,4,13], **Gitta H. Lubke**[1]

[1]Department of Psychology, University of Notre Dame, Notre Dame, IN; [2]Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA; [3]Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam; [4]Amsterdam Public Health research institute, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [5]Medical Research Council Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK; [6]Gillberg Neuropsychiatry Centre, University of Gothenburg, Gothenburg; [7]Centre for Ethics, Law and Mental Health (CELAM), University of Gothenburg, Gothenburg, Sweden; [8]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; [9]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [10]Department of Psychology, University of Jyvaskyla, Jyvaskyla, Finland; [11]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA; [12]Department of Public Health, University of Helsinki, Helsinki, Finland; [13]Amsterdam Neuroscience, Amsterdam, The Netherlands

Abstract

Background: Aggression in children has genetic and environmental causes. Studies of aggression can pool existing datasets to include more complex models of social effects. Such analyses require large datasets with harmonized outcome measures. Here, we made use of a reference panel for phenotype data to harmonize multiple aggression measures in school-aged children to jointly analyze data from five large twin cohorts.

Methods: Individual level aggression data on 86,559 children (42,468 twin pairs) were available in five European twin cohorts measured by different instruments. A phenotypic reference panel was collected which enabled a model-based phenotype harmonization approach. A bi-factor integration model in the integrative data analysis framework was developed to model aggression across studies while adjusting for rater, age, and sex. Finally, harmonized aggression scores were

Correspondence: Justin M. Luningham, Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA; jluningham@gsu.edu Dorret I. Boomsma, Department of Biological Psychology, VU University, Amsterdam, The Netherlands; di.boomsma@vu.nl.

analyzed to estimate contributions of genes, environment, and social interaction to aggression. The large sample size allowed adequate power to test for sibling interaction effects, with unique dynamics permitted for opposite-sex twins.

Results: The best-fitting model found a high level of overall heritability of aggression (~60%). Different heritability rates of aggression across sex were marginally significant, with heritability estimates in boys of ~64% and ~58% in girls. Sibling interaction effects were only significant in the opposite-sex twin pairs: the interaction effect of males on their female co-twin differed from the effect of females on their male co-twin. An aggressive female had a positive effect on male co-twin aggression, whereas more aggression in males had a negative influence on a female co-twin.

Conclusions: Opposite-sex twins displayed unique social dynamics of aggressive behaviors in a joint analysis of a large, multinational dataset. The integrative data analysis framework, applied in combination with a reference panel, has the potential to elucidate broad, generalizable results in the investigation of common psychological traits in children.

Introduction

Aggression in children is highly predictive of problems later in life, including delinquency, interpersonal relationship difficulties, depression, and difficulties in educational attainment, among other maladjustments (Boomsma, 2015; Brame, Nagin, & Tremblay, 2001; Fergusson, Horwood, & Ridder, 2005; Provencal, Booij, & Tremblay, 2015; Pulkkinen, 2017, 2018; Whipp et al., 2019). Clinical and subclinical levels of childhood aggression result in personal and societal costs to the children them-selves, their families, their peers and teachers, and their communities (Ettekal & Ladd, 2015; Foster, Jones, Bierman, & Coie, 2005). It is not clear if these detrimental results occur because of aggressive behavior itself, because aggression co-occurs with the majority of other internalizing and externalizing behavioral problems, or because of other common causes of negative outcomes and aggression (Bartels et al., 2018; see also www.action-euproject.eu). Data collected on twins may shed light on the etiology of aggression by parsing out the genetic and environmental contributions to childhood aggression. Collaboration among large-scale, multi-site twin studies allow for mega-analyses of twin data to achieve a broader understanding of the components of childhood aggression (Bennett et al., 2011; Budin-Ljøsne et al., 2013). However, phenotype harmonization across partnering studies presents a challenge for joint analyses. The current study bridges five studies across four countries using model-based phenotype harmonization and a *phenotypic reference panel* to investigate genetic and environmental components of childhood aggression in the Aggression in Children: Unraveling gene–environment interplay to inform Treatment and interventION strategies (ACTION) Consortium. Specifically, the current paper examines the sibling interaction effect that aggressive behavior of a child may have on the aggressive behavior of his or her co-twin.

Sibling interaction models, in which twin pairs directly influence one another, were first developed by Eaves (1976) and Carey (1986). These models posit an 'active' common source of phenotype expression in twins, rather than 'passive' similarities due only to shared genetic and environmental sources (Boomsma, 2014; Carey, 1986; Eaves, 1976; Eaves et al., 1978). A positive interaction coefficient captures sibling imitation, and negative interactions represent a departure between siblings due to their social dynamic. The exploration and quantification of such social interaction effects among twin pairs makes sense for behavioral traits such as childhood aggression, but is not often considered in behavior genetics, in part due to concerns of low statistical power. One study of aggression in Finnish twins aged 11 and 12 found a negative sibling interaction effect for same-sex male twin pairs, in which higher levels of aggression in one twin predicted lower levels in their co-twin (Vierikko, Pulkkinen, Kaprio, Viken, & Rose, 2003). Carey (1992) found evidence for sibling interaction in externalizing antisocial behavior (criminal registration) in Danish adults, and van den Oord, Boomsma, and Verhulst (1994) detected sibling effects for delinquency in models of a host of problem behaviors. The large, multinational European sample curated in ACTION provides an opportunity to test sibling interaction for overt aggression. There is evidence that sibling interaction effects are confounded with parental rater effects, indicating parents' perceived contrast of their children (Bartels et al., 2007; Simonoff et al., 1998). Because the data in ACTION are parent report, we must consider the possibility that sibling interaction effects may actually reflect rater contrast.

Meta-analytic reviews of aggression indicate that genetic components account for roughly 40%–60% of the variance in childhood aggression (Ferguson, 2010; Miles & Carey, 1997; Rhee & Waldman, 2002). These reviews survey extant results from numerous studies and glean conclusions from each of their individual results. One approach to forming a generalizable model of childhood aggression is to analyze twin data combined across multiple cohorts. The ACTION Consortium is a collaboration of large-scale, prospective longitudinal studies of childhood aggression, aiming to increase understanding of the biological, genetic, and environmental pathways of childhood aggression (Bartels et al., 2018; Boomsma, 2015).

One challenge facing ACTION (and many other multi-cohort behavioral studies) is measurement heterogeneity across studies. The large twin cohorts in ACTION used different instruments to measure aggression. If a harmonized phenotype can be achieved, the twin cohorts would provide a unique opportunity to fit a generalizable model of aggression to a large sample spanning several European countries that explores social interaction effects on top of genetic and environmental components.

In the current study, we utilize an integrative data analysis (IDA) framework for harmonization. IDA is an analytical method for modeling complex behavioral phenotypes across multiple independent studies (Curran & Hussong, 2009). IDA is an alternative to meta-analysis because the raw data are analyzed directly, and because psychometric models are used to link item-level data across studies and create comparable outcome scores for all subjects (Bauer & Hussong, 2009). IDA is also a departure from typical phenotype harmonization, where different questionnaire items are determined to be functionally the same item based on face validity, and differing item response categories are collapsed to the

same scale based on rational harmonization decisions (Gatz, Reynolds, Finkel, Hahn, & Zhou, 2015). This study uses an explicitly structured confirmatory factor analysis model fitted to all items to create a latent aggression score that is comparable across questionnaires.

The IDA harmonization approach allows for the direct modeling of some differences between cohorts, thus reducing phenotypic heterogeneity an increasing statistical power (Hussong et al., 2013; Luningham et al., 2019; van den Berg et al., 2014). However, overlapping survey information is needed across the different questionnaires in order to obtain a unified IDA model. One of the cohorts in ACTION recently collected data on several aggression questionnaires used by ACTION partners in a subsample of subjects. This *phenotypic reference panel* is essential to implement the IDA framework in a large consortium such as ACTION.

The objectives of the current paper are twofold. The first goal is to present a generalizable strategy for carrying out IDA of behavioral and psychological traits across multiple studies. This goal is addressed by detailing the IDA harmonization model developed for the aggression phenotype within the ACTION Consortium. The second goal is to ascertain effects of genotype, environment and social interaction on aggression and investigate sex differences.

Methods

Participants

School-aged children (ages 7–12) from five European cohorts in the ACTION Consortium were used in the current study, detailed below. All cohorts uploaded data about children's age, year of birth, gender, and twin zygosity. Twins were organized into five groups: monozygotic (MZ) males, dizygotic (DZ) males, MZ females, DZ females, and opposite-sex DZ (OSDZ) twins.

The Netherlands Twin Register (NTR).—Forthisstudy, mother and father reports for the NTR were used from collections when children were approximately age 7 (mean = 7.45, $SD = 0.40$), age 10 (mean = 9.94, $SD = 0.51$), and age 12 (mean = 12.21, $SD = 0.66$). The sample sizes of available data for mother-reported aggression scores were 24,780, 22,757, and 16,789 for ages 7, 10, and 12, respectively. The sample sizes for father report scores at the respective ages were approximately 17,430, 15,672, and 11,870. The sample was 50.2% female at age 7, 50.4% female at age 10, and 50.6% female at age 12. For details on data collection in the NTR, see, for example, van Beijsterveldt et al. (2013).

Twins Early Development Study (TEDS).—Parent report data for TEDS were included from collections when children were age 7 (mean = 7.06, $SD = 0.25$), age 9 (mean = 9.02, $SD = 0.29$), and age 11 (mean = 11.28, $SD = 0.70$). The data collection did not specify if respondents were fathers or mothers. The sample sizes at ages 7, 9, and 11 were approximately 15,668, 6,836, and 11,760, respectively. The sample was 51.3% female at age 7, 52.5% female at age 9, and 52.7% female at age 11. For details on TEDS, see Trouton, Spinath, and Plomin (2002).

Swedish Twin study of Child and Adolescent Development (TCHAD).—TCHAD collected data on approximately 2,200 children at age 8. Questionnaires were mailed to parents from the Swedish Twin Register. Exactage of the children was not available, and it was not specified which parent completed the report. The sample was 49.2% female. For details on TCHAD, see Lichtenstein, Tuvblad, Larsson, and Carlström (2007).

Child and Adolescent Twin Study in Sweden (CATSS).—Parents from the Swedish Twin Register were interviewed via telephone on the 9th birthday of their children, and follow-ups were conducted at age 12. Mother report data were available for approximately 18,495 children at age 9 and 5,737 children at age 12. Father report data were available for 3,362 children at age 9 and 730 children at age 12. The sample was approximately 50.5% female at age 9 and 51.5% female at age 12. For details on CATSS, see Anckarsäter et al. (2011).

FinnTwin12.—FinnTwin12 data were collected by mailing questionnaire assessments to families in the months before twins in the Finnish Twin Register turned 12 (mean age = 11.79, SD = 0.296). Responses were available from 2,724 families in total. Families were excluded if one or both co-twins had passed away or were living outside of Finland. Aggressive behavior was rated by the parents (60% rated by mothers alone, 37% by both parents together, 3% other). For details on FinnTwin12, see Kaprio, Pulkkinen, and Rose (2002).

The phenotypic reference panel.—The reference panel is a supplemental collection of participants from the NTR on three of the four questionnaires administered in the European cohorts. Throughout 2017, the complete CBCL and SDQ plus a selection of ATAC items were collected. Questionnaires were mailed to families with children around age 9 (mean = 9.42, SD = 0.78). The current study utilized mother report data on approximately 2,250 children and father report data on 1,540 children. The reference panel is 51.5% female.

Measures

Child Behavior Checklist (CBCL).—The CBCL 6–18 (Achenbach & Rescorla, 2001) was used by the NTR and TCHAD, as well as the reference panel. The CBCL 6–18 consists of 120 items which are rated on a 3-point scale ranging from 'not true = 0', 'somewhat or sometimes true = 1', to 'very true or often true = 2'. The CBCL 6–18 aggressive symptom subscale contains 18 original items, and 8 items pertaining to overt/physical aggression were retained based on factor analysis and overlap with other cohorts (see Lubke, McArtor, Boomsma, and Bartels (2018) for factor analysis of CBCL in NTR).

Strengths and Difficulties Questionnaire (SDQ).—The SDQ (Goodman, 1997) was administered in TEDS and the reference panel. The SDQ was designed to measure common mental health problems during childhood and adolescence. Ratings were on a three-point scale (with response options 'not true', 'somewhat true', 'certainly true'). The SDQ included five items on the conduct problems subscale that measured aggression.

Autism-Tics, ADHD, and other Comorbidities inventory (ATAC).—The ATAC (Larson et al., 2010) was administered in CATSS and the reference panel. The ATAC is a comprehensive screening interview for autism spectrum disorders, attention deficit/hyperactivity disorder, tic disorders, developmental coordination disorder, learning disorders, and other childhood mental disorders. The ATAC included 10 items related to aggression, and responses were scored on a 3-point scale (response options 'yes', 'yes, to some extent', and 'no').

Multidimensional Peer Nomination Inventory (MPNI).—MPNI was administered in FinnTwin12 and contains 37-items, creating three main domains with subscales and items as follows: (a) behavioral problems [direct proactive and reactive aggression (6 items), impulsivity-hyperactivity (7 items), inattention (4 items)]; (b) emotional problems [social anxiety (2 items), depression (5 items)], and (c) adjustment (12 items). For this analysis, 4 items from direct aggression plus one item from hyperactivity-impulsivity subscale were utilized to represent aggressive behaviors. For each question, parents rated how well the description of the item fit the twin in question on a scale from 0 (does not fit the child at all) to 3 (the characteristic fits the child very well) (Pulkkinen et al., 1999). Although the MPNI was not administered in the reference panel, items were worded similarly enough to the other questionnaires to include this questionnaire.

Phenotypic reference panel

The factor analytic method used to carry out IDA explicitly models the covariances among all items used across the combined cohorts. When items are unique to one and only one cohort, the covariance matrix between those items and items among other cohorts are completely missing, resulting in zero covariance coverage where items do not overlap. This precludes the use of many frequentist methods for estimating factor models and necessitates very strong assumptions for Bayesian estimation. The phenotypic reference panel was collected to overcome this limitation by providing links across items that were otherwise collected in only one of the cohorts (e.g., SDQ and ATAC) and across subjects that responded to otherwise unique items (e.g. the reference panel and TEDS for SDQ items).

Measurement model IDA strategy

Initially, items corresponding to a unidimensional subtype of overt/physical aggression were identified based on a combination of face validity assessment and existing literature. Because the MPNI items were missing in the reference set, the MPNI items were harmonized to existing items from the other three questionnaires based on face validity and on similar response frequencies to matched items. Two MPNI items for proactive aggression (MPNI25 and MPNI13) and one item for disobedience (MPNI33) were direct matches in wording, while two items for reactive aggression (MPNI21 and MPNI27) were matches with items for very low self-regulation. Items selected for aggression are presented in Table S1. Note that this reflects the more commonly used rational harmonization approach; however, we corrected for the potential impact of rational harmonization within our harmonization model (see below).

Next, an integration model incorporating measurements from ages 7 to 12 and both mother and father reported data were fitted. For this analysis, TEDS, TCHAD, and FinnTwin12 data were treated as mother-reported data, because in general, mothers were more likely to participate in surveys about their children than fathers. We employed an adaptation of the multi-rater model described by Bauer et al. (2013). In this model, all of the items load onto a single factor, representing the aggression score. In other words, the shared variance among all items can be represented as the underlying aggression factor, given that all items are intended to tap into aspects of aggression. Beyond the general aggression factor common to all items, there were two sources of additional covariance among item subsets: rater effects (mother and father items for each child) and item effects (the same identical question was administered to both parents for each child). Therefore, the model included two types of specific factors: rater factors and residual factors. All CBCL items and SDQ items 07 and 12 loaded onto either mother or father factor, depending on the rater. The ATAC and SDQ05 items had negative residuals when loading on the rater factors, indicating overfitting, so these loadings were excluded. A residual factor was modeled for each pair of the same items administered to the two raters, excluding ATAC63, SDQ05, and SDQ12 (models with these factors again indicated overfitting). Age was regressed out of the items to adjust for slight age differences across cohort, and a cohort dummy variable was also regressed out of items from the FinnTwin12 cohort, due to potential item differences in the harmonized MPNI items. Finally, additional residual correlations were included for items across questionnaires with extremely similar wording: the 'anger' items ATAC63, SDQ05, and CBCL095, and 'destroy' items CBCL020 and CBCL021.

The model was fitted as a multi-group model with separate parameters for boys and girls due to expected gender differences in overt aggression (Card, Stucky, Sawalani, & Little, 2008). The integration model was evaluated on convergence, model fit, and correlations of factor scores with sum scores in the cohorts. A schematic of the final integration model is presented in Figure 1.

Integrated twin modeling

Twin models were fitted to the harmonized aggression phenotype to evaluate genetic and environmental contributions to aggression, gender differences in aggression, and potential interactions between co-twins. Twin models can partition the variance of a phenotype into components due to additive genetic effects (A), common environment (C), and nonshared environment (E). The twin-modeling framework estimates these variance components by modeling the expected relationships between pairs of twins (Neale & Cardon, 1992; Rijsdijk & Sham, 2002). For example, MZ twins are expected to be genetically identical, but DZ twins share on average 50% of their segregating genetic information. A multi-group model is fitted to pairs of MZ twins and DZ twins, in which the A component is perfectly correlated among MZ twins and correlated 0.5 for DZ twins. The C component is perfectly correlated across twins of both zygosity groups to capture shared variance not found in the A component, and the unique component E is completely uncorrelated across twin pairs. The E component includes all nonshared environmental effects as well as measurement error. The variance component factors are assumed to be standard normally distributed. Sex-limitation models further allow for the modeling of sex differences when male, female, and opposite-

sex twin pairs are present. The common effects sex-limitation model allows the path coefficients to vary across male and female groups, testing if the same underlying genetic and environmental components are at work for both males and females but the magnitudes differ. For the OSDZ group, the path coefficients are constrained to equal the path coefficients of male and female same-sex pairs for model identification. Finally, qualitative sex differences were modeled by permitting the correlation between genetic components to be freely estimated in the OSDZ twins rather than fixed at 0.5 (for other examples of these models, see Eley, Lichtenstein, & Stevenson, 1999; Happonen et al., 2002).

Reciprocal relationships across siblings can be modeled by allowing for direct effects from one twin's phenotype to the other. A sibling interaction is often indicated by differences in the variances across twin groups (Carey, 1986). The presence of the sibling direct effect terms changes the expected variance attributable to each of the A, C, and E components across MZ and DZ twins within each sex (Appendix S1; see table 1 of Rietveld et al., 2003). The interaction can be fixed across all zygosity groups, or it can be freely estimated in all five groups. In the current analysis, the potential for sibling interaction was evaluated after twin group differences and sex-limitation was established.

Analysis plan

Twin correlations were first estimated in a saturated model to evaluate mean and variance differences in overt aggression across the different twin groups and across sex. The ACE model was fitted and compared to the more parsimonious alternatives, such as the AE and CE models. Sex differences were analyzed through sex-limitation models. The common effects sex-limitation model was then fitted within the best decomposition model. Models allowing for differences across sex in both variance components and means were compared. The general sex-limitation model, allowing for the genetic correlation in the OSDZ group to be < 0.5, was also estimated.

As a final step, sibling interaction paths were tested. A sibling interaction parameter was initially constrained across all five groups and was then allowed to vary across sexes, both without (two interaction parameters) and with (three interaction parameters) a separate parameter for the OSDZ group. As an exploratory extension of more typical interaction models, a model with four direct effect terms was included – an interaction for same-sex males, same-sex females, a male twin effect on his female co-twin, and a female twin effect on her male co-twin. This model allows for a unique social dynamic for the OSDZ twins relative to their same-sex counterparts. Models were compared based on nested χ^2 difference tests, along with model fit criteria (i.e., root mean square error of approximation (RMSEA), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Models were fitted in Mplus version 7 (Muthén & Muthén, 1998).

Results

Descriptive statistics

Sample counts were obtained after the data were cleaned for completeness within each cohort and after repeated measurements were removed. In case a given study provided

repeated measurements, the closest available observation to age 9.5 was retained. Each cohort group had a fairly even representation of both boys and girls. The response proportions for each overt aggression item are presented in Table S2.

Integrated factor scores

In total, 86,559 individuals were available in the combined data, with 42,468 complete twin pairs.[1] There was an even distribution of MZ, DZ, and OSDZ twins. Sample sizes for each cohort and each zygosity group are presented in Tables S3 and S4, respectively. Factor scores were saved from the multi-group, multi-rater IDA model. The model converged and displayed good overall fit to the data, with a root mean square error of approximation (RMSEA) of 0.008 and a Comparative Fit Index (CFI) of 0.98. The factor loadings of the model are presented in Table S5. The overall factor score means and cohort-and sex-specific means are presented in Table S6. Factor score means varied slightly across cohort, but no cohort displayed means that were significantly different from zero (see large standard deviations relative to the means in the Table S6). Boys tended to have higher factor score means than girls, except for similar scores in CATSS. The IDA model was refitted without the 4,884 samples from the FinnTwin12 study, due to the fact that the MPNI items from FinnTwin12 were included in the overall dataset based on face validity and item response distributions. The correlation between the harmonized aggression scores for the remaining subjects across the two models was 0.99 ($SE = 0.001$).

Twin modeling

As a first step, phenotypic variances, covariances, and means in twin pairs across the different zygosity groups were obtained from a saturated model, presented in Table 1. The MZ twins had larger correlations than DZ twins, indicating that a genetic component was present. The DZ correlations were more than half the MZ correlations, meaning that a common environment component was likely needed to account for phenotypic variance. The same-sex DZ males had slightly larger variances than MZ males, and same-sex DZ females had notably larger variances than MZ females, indicating a possible sibling interaction. The OSDZ correlation was smaller than the same-sex DZ correlations, indicating there were sex-specific genetic and/or environmental effects at play, or that the OSDZ group was otherwise meaningfully different from the same-sex twin pairs. There were clear differences in mean levels of aggression for males and females, presented in Table 2.

Table 3 presents the model fit information for all models, including the χ^2 statistic, degrees of freedom, RMSEA, AIC, and BIC. The constrained ACE model, with path coefficients set equivalent across all 5 groups, fit better than the AE and CE model, indicating that the A, C, and E components were all appropriate for these data. Means were estimated separately for males and females due to significant differences in the saturated model.

---

[1]·There is a discrepancy between the total number of individuals used in the IDA model vs the twin model for two reasons. First, 579 twin pairs had missing or mis-matched zygosity codes, so were excluded from the twin analysis but used in the IDA model for aggression. Secondly, there were roughly 500 individuals that were from triplets, quadruplets, etc. All individuals were used for the IDA portion, but only twin pairs were used for the twin model.

The common effects sex-limitation model was then fitted, providing a large improvement in fit relative to the ACE only model, $\chi^2(3) = 747.97$, $p < .001$. In this model, the constraints on the path coefficients were relaxed to correspond to either male or female twins, with sex-specific constraints applied to opposite-sex twins as well. Based on results from the five-group saturated model, the OSDZ group was allowed to have its own sex-specific means, resulting again in significant improvement to the model ($\chi^2(2) = 59.67$, $p < .001$). Finally, freely estimating the genetic correlation in the OSDZ group slightly improved model fit ($\chi^2(1) = 4.3$, $p = .04$), but this genetic correlation coefficient was very near 0.5 (0.46, $SE = 0.02$).

The final series of models explored sibling interaction effects. Adding a single interaction, constrained across all twin groups, did not significantly improve model fit over the sex-limitation model. However, allowing a sex-specific interaction did improve model fit, $\chi^2(2) = 15.18$, $p < .001$. A model with sex-specific interaction for same-sex twins plus two unique sibling effects in the OSDZ group resulted in a larger improvement in fit, $\chi^2(2) = 23.85$, $p < .001$. A path diagram detailing the parameters included in the final model is presented in Figure 2.

The parameter estimates for the best-fitting model are presented in Table 4. The interaction effect for same-sex twins was small and nonsignificant. The interaction terms for the OSDZ group were significant and in reciprocal directions, yielding results with potentially interesting interpretations. In this model, a higher level of aggression in a male twin results in less aggression in a female co-twin. However, females with higher aggression tend to interact with their male co-twin by promoting higher aggression. The fact that this model provided the best overall fit offered some evidence that the OS twins indeed have a unique dynamic that differs from same-sex twin pairs. The proportion of variance attributable to each variance component is presented in Table 5. The variance components were different for OSDZ twins due to the male- and female-specific interaction terms. In males, aggression was about 64% heritable. The common environmental component contributed about 8%, and unique environment contributed 28%. Aggression was estimated to be 58% heritable in females, but this difference was only marginally significantly less than the estimate for males (difference = 0.059, $SE = 0.038$, one-tailed $p = .06$). The common environmental component contributed 13%, and the unique environment contributed 29%. For the OS twins, the contribution of each component varied across male and female twins due to the significant sibling interaction coefficient, see Appendix S1. The partitioning of genetic and environmental variance, however, was very much in line with twins from same-sex pairs. Therefore, the OSDZ pairs had a direct influence on each other that differed from same-sex pairs, but the resultant heritability estimates of aggression were fairly consistent for all males and all females.

Discussion

The twin analyses presented here introduced the innovative use of integrated factor scores as a harmonized phenotype across multiple large cohorts, resulting in a large twin model that found a unique dynamic among OSDZ twins. The heritability estimates discovered in this analysis were consistent with existing literature finding heritability in boys around school

age to be typically around 55%–67% and heritability in girls to be a bit lower, around 45%–58% (van Beijsterveldt et al., 2003; Porsch et al., 2016).

The sibling interaction component was detected in the OSDZ group, adding a more explicit interpretation to the dynamics of those twins' shared environment. The interaction effect in the OSDZ group signals that OSDZ twins have a special social learning context, as demonstrated by Pulkkinen et al. (2003). This study showed that the OSDZ girls compared to singletons were more aggressive, which is typical of boys, while the OSDZ boys compared to singletons were more prosocial (compliant and constructive), which is typical of girls. However, those results were based on peer nomination; in the current study, results are based on parent ratings, explaining some differences in findings. Our final model also accounted for the social learning that may occur between OSDZ twins. Parents that perceived high levels of aggression in male OS twins rated the female co-twin as less aggressive. The small but positive effect of females on their male co-twins, conversely, indicated that aggression in a female OS twin in some way triggered aggression in the male co-twin, or that parents were more likely to perceive aggression in the male co-twin.

There are two important notes of caution worth mentioning when interpreting the twin interaction result. Firstly, there is evidence that sibling interaction effects are confounded with rater effects (Bartels et al., 2007; Simonoff et al., 1998). These data are strictly parental perceptions of their children's behavior, so the sibling contrast effect in OSDZ twins could reflect rater bias in how parents view aggressive behaviors of their male and female children. However, because we only detected the sibling interaction in the OSDZ group, this may represent a true social dynamic rather than a contrast effect. Secondly, allowing for sex-specific interaction effects in the OSDZ group may simply indicate that these effects are capturing additional sex differences instead of a direct dynamic specific to OS twins. The superior model fit of the sibling interaction model compared to other sex-difference dynamics provides some confidence in the interpretation.

Another unique component of this analysis was the use of phenotype scores harmonized across multiple studies. The current analysis was able to make a more generalized statement about aggression in school-aged children from the Netherlands, Finland, the UK, and Sweden. Combining data across cohorts was also beneficial for these data because of the low prevalence of overt aggression behaviors. If these data were analyzed separately, and the data were further split into the five zygosity groups for twin modeling, there would potentially be sparse responses at the item level and limited variability in the observed scores. The collection of the reference panel was crucial for modeling all items across cohorts. Further, the integrated aggression score in school-aged children is available to ACTION partners for future joint analyses, and the reference panel can facilitate future linking and scaling of phenotype scores across the consortium. As with many psychological traits, indicators of aggressive behavior are generally diverse across studies. Multi-study collaborations or comparisons across studies can benefit from item construct analyses and the conceptual study of the construct that is being measured.

This integrated score was not without limitations. The large sample size and relatively large number of items across cohort and rater mean that increasingly complex integration models

could be fitted. We limited our scope of integration models to those that made sense conceptually, given our interest in overt aggression. We did not carry out formal tests of model comparison among numerous competing models, however. Our model accounted for cohort differences by regressing out an indicator for FinnTwin12 samples, because these samples used items that were not included in the reference panel and were incorporated based on face validity and rational harmonization of similar response rates. Future integration models could account for additional cohort effects more explicitly (by regressing item parameters on cohort indicators, see Curran et al., 2014), or cohort membership could be controlled for in twin models directly. Still, the benefit of the integrated analysis was in creating a large sample with a single comparable score. This ensured that we had large representation of each zygosity group, and we were able to fit more complex twin models with social dynamics unique to OSDZ twins in the form of sibling interaction parameters.

The IDA framework discussed here provides promise for multi-study collaborations of behavioral or psychological phenotypes. Utilizing a phenotypic reference panel and any overlapping item content, researchers can create scaled scores for the phenotype that are adjusted in part to account for at least some of the differences between cohorts, thus reducing phenotypic heterogeneity (Curran & Hussong, 2009). This is advantageous especially given the fact that the trait of interest is often measured using different scales/ instruments in partnering studies. Ultimately, the same integrated phenotype scores used here can be used not only in twin models but also any type of joint analysis. For genetic consortia like ACTION, these include genome-wide association studies and epigenetic models investigating aggression. To the extent that item content varies across studies, a psychometric model in the IDA framework should remove this specific type of measurement heterogeneity and benefit the joint analyses of multiple studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Achenbach TM, & Rescorla LA (2001). Empirically based and dsm-oriented assessment of preschoolers for pharma-cotherapy and other interventions. Child and Adolescent Psychopharmacology News, 6, 1–7.

Anckarsäter H, Lundström S, Kollberg L, Kerekes N, Palm C, Carlström E, … & Lichtenstein P (2011). The Child and Adolescent Twin Study in Sweden (CATSS). Twin Research and Human Genetics, 14, 495–508. [PubMed: 22506305]

Bartels M, Boomsma DI, Hudziak JJ, van Beijsterveldt TCEM, & van den Oord EJCG (2007). Twins and the study of rater (dis)agreement. Psychological Methods, 12, 451–466. [PubMed: 18179355]

Bartels M, Hendriks A, Mauri M, Krapohl E, Whipp A, Bolhuis K, … & Boomsma DI (2018). Childhood aggression and the co-occurrence of behavioural and emotional problems: results across ages 3–16 years from multiple raters in six cohorts in the EU-ACTION project. European Child and Adolescent Psychiatry, 27, 1105–1121. [PubMed: 29845340]

Bauer DJ, Howard AL, Baldasaro RE, Curran PJ, Hussong AM, Chassin L, & Zucker RA (2013). A trifactor model for integrating ratings across multiple informants. Psychological Methods, 18, 475–493. [PubMed: 24079932]

Bauer DJ, & Hussong AM (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. Psychological Methods, 14, 101–125. [PubMed: 19485624]

Bennett S, Caporaso N, Fitzpatrick AL, Agrawal A, Barnes K, Boyd H, … & Williams K (2011). Phenotype harmonization and cross-study collaboration in GWAS consortia: The Geneva experience. Genetic Epidemiology, 35, 159–173. [PubMed: 21284036]

Boomsma DI (2014). Sibling interaction effects In Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL (Eds.), Wiley StatsRef: Statistics Reference Online (pp. 1–3). Chichester: John Wiley & Sons Ltd.

Boomsma DI (2015). Aggression in children: unravelling the interplay of genes and environment through (epi)genetics and metabolomics. Journal of Pediatric and Neonatal Individualized Medicine, 4, e040251.

Brame B, Nagin DS, & Tremblay RE (2001). Developmental trajectories of physical aggression from school entry to late adolescence. Journal of Child Psychology and Psychiatry, 42, 503–512. [PubMed: 11383966]

Budin-Ljøsne I, Isaeva J, Knoppers BM, Tassé AM, Shen HY, McCarthy MI, … & ENGAGE Consortium (2013). Data sharing in large research consortia: experiences and recommendations from ENGAGE. European Journal of Human Genetics, 22, 317–321. [PubMed: 23778872]

Card NA, Stucky BD, Sawalani GM, & Little TD (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. Child Development, 79, 1185–1229. [PubMed: 18826521]

Carey G (1986). Sibling imitation and contrast effects. Behavior Genetics, 16, 319–341. [PubMed: 3753368]

Carey G (1992). Twin imitation for antisocial behavior: Implications for genetic and family environment research. Journal of Abnormal Psychology, 101, 18–25. [PubMed: 1537964]

Curran PJ, & Hussong AM (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. Psychological Methods, 14, 81–100. [PubMed: 19485623]

Curran P, Mcginley J, Bauer D, Hussong A, Burns A, Chassin L, … Zucker R (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. Multivariate Behavioral Research, 49, 214–231. [PubMed: 25960575]

Eaves LJ (1976). A model for sibling effects in man. Heredity, 36, 205–20514. [PubMed: 1063734]

Eaves LJ, Last KA, Young PA, & Martin NG (1978). Model-fitting approaches to the analysis of human behaviour. Heredity, 41, 249–320. [PubMed: 370072]

Eley TC, Lichtenstein P, & Stevenson J (1999). Sex Differences in the etiology of aggressive and nonaggressive antisocial behavior: Results from two twin studies. Child Development, 70, 155–168. [PubMed: 10191520]

Ettekal I, & Ladd GW (2015). Costs and benefits of children's physical and relational aggression trajectories on peer rejection, acceptance, and friendships: Variations by aggression subtypes, gender, and age. Developmental Psychology, 51, 1756–1770. [PubMed: 26414097]

Ferguson CJ (2010). Genetic contributions to antisocial personality and behavior: A meta-analytic review from an evolutionary perspective. Journal of Social Psychology, 150, 160–180. [PubMed: 20397592]

Fergusson DM, John Horwood L, & Ridder EM (2005). Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. Journal of child psychology and psychiatry, 46, 837–849. [PubMed: 16033632]

Foster E, Jones D, Bierman K, & Coie J (2005). The high costs of aggression: Public expenditures resulting from conduct disorder. American Journal of Public Health, 95, 1767–72. [PubMed: 16131639]

Gatz M, Reynolds CA, Finkel D, Hahn CJ, Zhou Y, & Zavala C (2015). Data harmonization in aging research: Not so fast. Experimental Aging Research, 41, 475–495. [PubMed: 26524232]

Goodman R (1997). The strengths and difficulties questionnaire: A research note. Journal of Child Psychology and Psychiatry, 38, 581–586. [PubMed: 9255702]

Happonen M, Pulkkinen L, Kaprio J, Van der Meere J, Viken RJ, & Rose RJ (2002). The heritability of depressive symptoms: multiple informants and multiple measures. Journal of Child Psychology and Psychiatry, 43, 471–479. [PubMed: 12030593]

Hussong AM, Curran PJ, & Bauer DJ (2013). Integrative data analysis in clinical psychology research. Annual Review of Clinical Psychology, 9, 61–89.

Kaprio J, Pulkkinen L, & Rose R (2002). Genetic and environmental factors in health-related behaviors: Studies on Finnish Twins and Twin families. Twin Research, 5(5), 366–371. [PubMed: 12537860]

Larson T, Anckarsäter H, Gillberg C, Ståhlberg O, Carlström E, Kadesjö B, …& Gillberg C (2010). The autism–tics, AD/HD and other comorbidities inventory (A-TAC): Further validation of a telephone interview for epidemiological research. BMC Psychiatry, 10, 1. [PubMed: 20055988]

Lichtenstein P, Tuvblad C, Larsson H, & Carlström E (2007). The Swedish twin study of CHild and adolescent development: The TCHAD-study. Twin Research and Human Genetics, 10, 67–73. [PubMed: 17539366]

Lubke GH, McArtor DB, Boomsma DI, & Bartels M (2018). Genetic and environmental contributions to the development of childhood aggression. Developmental psychology, 54, 39–50. [PubMed: 29058931]

Luningham JM, McArtor DB, Hendricks AH, van Beijsterveldt CEM, Lichtenstein P, Lundström S, … & Lubke GH (2019). Data integration methods for phenotype harmonization in multi-cohort genome-wide association studies with behavioral outcomes. Frontiers in Genetics, 10, 1227. [PubMed: 31921287]

Miles DR, & Carey G (1997). Genetic and environmental architecture on human aggression. Journal of personality and social psychology, 72, 207. [PubMed: 9008382]

Muthén LK, & Muthén BO (1998–2015). Mplus user's guide (7th edn). Los Angeles: Author.

Neale MC, & Cardon LR (1992). Methodology for genetic studies of twins and families. Doordrecht, the Netherlands: Kluwer Academic.

Porsch RM, Middeldorp CM, Cherny SS, Krapohl E, van Beijsterveldt CEM, Loukola A, … & Bartels M (2016). Longitudinal heritability of childhood aggression. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 171, 697–707. [PubMed: 26786601]

Provencal N, Booij L, & Tremblay R (2015). The developmental origins of chronic physical aggression: Biological pathways triggered by early life adversity. Journal of Experimental Biology, 218, 123–133. [PubMed: 25568459]

Pulkkinen L (2017). Human development from middle childhood to middle adulthood: Growing up to be middle-aged (In collaboration with Katja Kokko). London: Routledge.

Pulkkinen L (2018). Longitudinal Study of Personality and Social Development: Insights about aggression after five decades In Vazsonyi AT, Flannery DJ & DeLisi M (Eds.), The Cambridge handbook of violent behavior and aggression (2nd edn., pp. 31–51). Cambridge, UK: Cambridge University Press.

Pulkkinen L, Kaprio J, & Rose RJ (1999). Peers, teachers and parents as assessors of the behavioural and emotional problems of twins and their adjustment: The multidimensional peer nomination inventory. Twin Research and Human Genetics, 2, 274–285.

Pulkkinen L, Vaalamo I, Hietala R, Kaprio J, & Rose RJ (2003). Peer reports of adaptive behavior in twins and singletons: Is twinship a risk or an advantage? Twin Research, 6, 106–118. [PubMed: 12723997]

Rhee SH, & Waldman ID (2002). Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. Psychological Bulletin, 128, 490–529. [PubMed: 12002699]

Rietveld M, Posthuma D, Dolan C, & Boomsma DI (2003). Adhd: Sibling interaction or dominance: An evaluation of statistical power. Behavior Genetics, 33, 247–255. [PubMed: 12837015]

Rijsdijk V, & Sham PC (2002). Analytic approaches to twin data using structural equation models. Briefings in Bioinformatics, 3, 119–133. [PubMed: 12139432]

Simonoff E, Pickles A, Hervas A, Silberg JL, Rutter M, & Eaves LJ (1998). Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. Psychological Methods, 28, 825–837.

Trouton A, Spinath F, & Plomin R (2002). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. Twin Research, 5, 444–448. [PubMed: 12537874]

van Beijsterveldt CEM, Bartels M, Hudziak JJ, & Boomsma DI (2003). Causes of stability of aggression from early childhood to adolescence: A longitudinal genetic analysis in Dutch twins. Behavior Genetics, 33, 591–605. [PubMed: 14574135]

van Beijsterveldt C, Groen-Blokhuis M, Hottenga J, Frani S, Hudziak J, Lamb D, …& Boomsma D (2013). The young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. Twin Research and Human Genetics, 16, 252–267. [PubMed: 23186620]

van den Berg SM, de Moor MHM, McGue M, Pettersson E, Terracciano A, Verweij KJH, … & Boomsma DI (2014). Harmonization of neuroticism and extraversion phenotypes across inventories and cohorts in the genetics of personality consortium: An application of item response theory. Behavior Genetics, 44, 295–313. [PubMed: 24828478]

van den Oord EJ, Boomsma DI, & Verhulst FC (1994). A study of problem behaviors in 10- to 15-year-old biologically related and unrelated international adoptees. Behavior Genetics, 24, 193–205. [PubMed: 7945150]

Vierikko E, Pulkkinen L, Kaprio J, Viken R, & Rose R (2003). Sex differences in genetic and environmental effects on aggression. Aggressive Behavior, 29, 55–68.

Whipp AM, Korhonen T, Raevuori A, Heikkilä K, Pulkkinen L, Rose R, … & Vuoksimaa E (2019). Early adolescent aggression predicts antisocial personality disorder in young adults: A population-based study. European Child and Adolescent Psychiatry, 28, 341–350. [PubMed: 30019148]
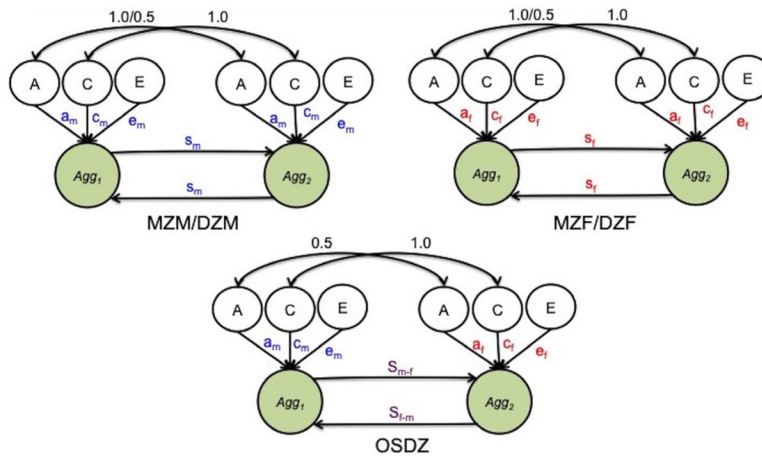
Author Manuscript

Key points

- Overt aggression is a moderately heritable problem behavior in children. Heritability estimates vary across different studies based on different European twin registers.

- Integrative data analysis was used to formulate a model-based harmonization approach for creating a comparable aggression score across multiple cohorts in the ACTION Consortium.

- Twin models of harmonized aggression resulted in relatively high estimates of heritability of aggression (60% overall), with marginally significant sex differences (64% for males, 58% for females). In addition, a unique sibling contrast dynamic was detected in opposite-sex twin pairs.

- The large sample size from combining studies provided statistical power to detect these sex differences.

- Integrative data analysis provides a framework for cross-study collaboration with difficult-to-measure psychological outcomes.
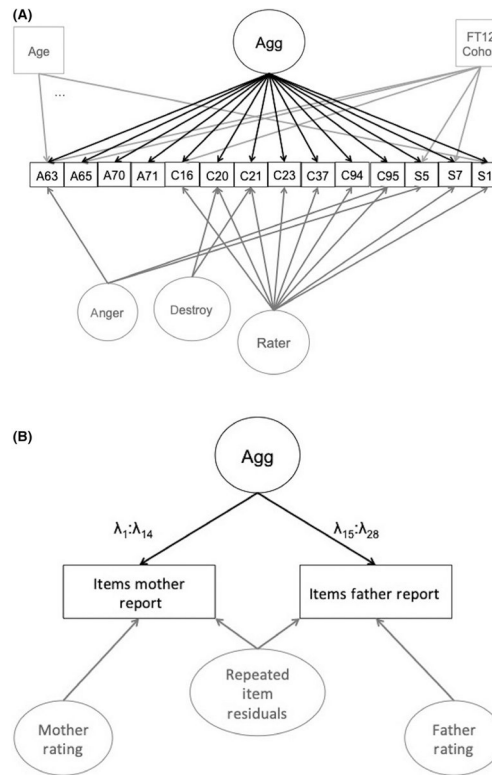
Figure 1.
Path diagram of an adapted multi-rater bi-factor model fitted to data across three
questionnaires, two raters, and six cohorts. The key paths of interest are in black. Gray paths
are estimated, but are of secondary interest, accounting for additional sources of covariance
(rater, item residuals) or adjusting for covariates (age, FinnTwinn12 cohort). λ represent the
factor loadings, and the different subscripts of the items loading on aggression for mother
and father items indicate that they have different weights. Repeated item residuals (e.g., a
residual between mother report CBCL16 and father report CBCL16) have fixed loadings
across rater. *Note*: Agg = aggression; A = ATAC; C = CBCL; S = SDQ; FT12 = FinnTwin12

Figure 2.
Path diagram of common effects sex-limitation model with unique 'c' and 'e' paths for the OSDZ group, with unique sibling interactions added. NOTE: colors correspond to model parameters that vary across groups

Table 1

Phenotypic variances, covariances, and correlations in the saturated twin model

	MZM		DZM		MZF		DZF		OSD*Z*	
	Agg1	**Agg2**	**Agg1**	**Agg2**	**Agg1**	**Agg2**	**Agg1**	**Agg2**	**Agg1**	**Agg2**
Agg1	0.409	(0.736)	0.421	(0.433)	0.318	(0.693)	0.328	(0.402)	0.387	(0.366)
Agg2	0.294	0.389	0.175	0.389	0.211	0.292	0.129	0.314	0.121	0.281

Agg1 is the score for twin 1, and Agg2 is the score for twin 2. For OSDZ twins, the male twin is always twin 1 and the female twin is always twin 2. Correlations presented in parentheses.

Table 2

Mean aggression scores from the saturated twin model

	MZM		DZM		MZF		DZF		OSDZ	
	Agg1	**Agg2**	**Agg1**	**Agg2**	**Agg1**	**Agg2**	**Agg1**	**Agg2**	**Agg1**	**Agg2**
Mean	0.305	0.277	0.293	0.270	0.108	0.083	0.127	0.100	0.243	0.068
SD	0.008	0.008	0.008	0.007	0.007	0.006	0.007	0.007	0.005	0.004

Agg1 is the score for twin 1, and Agg2 is the score for twin 2. For OSDZ twins, the male twin is always twin 1 and the female twin is always twin 2. For the current analysis, twin 1 and twin 2 codes were arbitrarily assigned for same-sex pairs.

Table 3

Model fit information for twin variance decomposition models

Model	Par.	χ^2	df	χ^2 (DF)	RMSEA	AIC	BIC
CE model	4	36,580.02	21	-	0.143	139,257.78	139,292.70
AE model	4	1,048.62	21	-	0.076	136,648.38	136,683.01
ACE	5	1,004.24	20	44.38(1) ***	0.076	136,605.99	136,649.28
ACE sex-limitation	8	256.27	17	747.96(3) ***	0.041	135,864.04	135,933.29
ACE sex-limit. with additional free means in OSDZ group	10	196.60	15	59.67(2) ***	0.038	135,808.36	135,894.93
ACE sex-limit. with estimated genetic correlation in OSDZ	11	192.25	14	4.35(1) *	0.039	135,806.01	135,901.26
ACE sex-limit. + interaction	11	193.11	14	3.49(1) †	0.039	135,806.88	135,902.10
ACE sex-limit. + sex-specific interactions	12	181.42	13	11.69(1) ***	0.039	135,797.19	135,901.06
ACE sex-limit + 4 interactions: male->male, female- >female, male->female, female->male	14	157.57	11	23.85(2) ***	0.040	135,777.33	135,898.53

Lower values of RMSEA, AIC, and BIC indicate better model fit. refers to the change in χ^2 and *df* compared to the previous model. The change in χ^2 for nested models follows a χ^2 distribution with *df* equal to the change in *df*, with significant differences indicating significant improvement in model fit. *df*, degrees of freedom; Par., parameter.

†
$p = .062$

*
$p = .037$

***
$p < .001$

Table 4

Parameter estimates of the best-fitting model

	Par.	**Est.**	**SE**	**p**
Male	a_m	0.493	0.010	<.001
	c_m	0.175	0.054	.001
	e_m	0.333	0.007	<.001
	s_m	0.023	0.018	.200
Female	a_f	0.428	0.011	<.001
	c_f	0.200	0.050	<.001
	e_f	0.303	0.008	<.001
	s_f	−0.014	0.024	.569
OS	$s_{m\text{-}f}$	−0.092	0.026	<.001
	$s_{f\text{-}m}$	0.069	0.034	.041
Intercepts	$\beta_{0,m}$	0.279	0.007	<.001
	$\beta_{0,f}$	0.106	0.005	<.001
	$\beta_{0,m,os}$	0.239	0.006	<.001
	$\beta_{0,f,os}$	0.90	0.008	<.001

Est., estimate; *p*, *p*-value; Par., parameter; *SE*, standard error. The intercept subscripts 'm', 'f', 'm,os' and 'f,os' refer to intercept terms from same-sex male, same-sex female, male in the OS group, and female in the OS group, respectively. The presented path coefficients are unstandardized; see Table 5 for proportional variance components and Appendix S1 for notes on how they were obtained.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Proportional Variance Contribution (and C.I.) of Each Component for Male, Female, and OSDZ Groups

Male		Female		OSDZ	
Comp.	Prop. Variance	Comp.	Prop. Variance	Comp.	Prop. Variance
A	.64 [.56, .72]	A	.58 [.49, .67]	Male A	.64 [.56, .72]
C	.08 [0, .17]	C	.13 [.02, .24]	Female A	.57 [0.50, .64]
E	.28 [.26, .30]	E	.29 [.26, .32]	Male C	.09 [0, .18]
				Female C	.11 [.01, .22]
				Male E	.27 [.25, .30]
				Female E	.32 [.28, .35]

Comp., component; C.I., confidence interval; Prop., proportional.