Behavior Genetics
https://doi.org/10.1007/s10519-021-10044-0

ORIGINAL RESEARCH

The Augmented Classical Twin Design: Incorporating Genome-Wide Identity by Descent Sharing Into Twin Studies in Order to Model Violations of the Equal Environments Assumption

Liang-Dar Hwang[1,2] · Brittany L. Mitchell[2,3] · Sarah E. Medland[2] · Nicholas G. Martin[2,3] · Michael C. Neale[4] · David M. Evans[1,2,5,6]

Received: 25 August 2020 / Accepted: 21 January 2021

Abstract

The Classical Twin Method (CTM) compares the similarity of monozygotic (MZ) twins with that of dizygotic (DZ) twins to make inferences about the relative importance of genes and environment in the etiology of individual differences. The design has been applied to thousands of traits across the biomedical, behavioral and social sciences and is arguably the most widely used natural experiment known to science. The fundamental assumption of the CTM is that trait relevant environmental covariation within MZ pairs is the same as that found within DZ pairs, so that zygosity differences in within-pair variance must be due to genetic factors uncontaminated by the environment. This equal environments assumption (EEA) has been, and still is hotly contested, and has been mentioned as a possible contributing factor to the missing heritability conundrum. In this manuscript, we introduce a new model for testing the EEA, which we call the Augmented Classical Twin Design which uses identity by descent (IBD) sharing between DZ twin pairs to estimate separate environmental variance components for MZ and DZ twin pairs, and provides a test of whether these are equal. We show through simulation that given large samples of DZ twin pairs, the model provides unbiased estimates of variance components and valid tests of the EEA under strong assumptions (e.g. no epistatic variance, IBD sharing in DZ twins estimated accurately etc.) which may not hold in reality. Sample sizes in excess of 50,000 DZ twin pairs with genome-wide genetic data are likely to be required in order to detect substantial violations of the EEA with moderate power. Consequently, we recommend that the Augmented Classical Twin Design only be applied to datasets with very large numbers of DZ twin pairs (> 50,000 DZ twin pairs), and given the strong assumptions relating to the absence of epistatic variance, appropriate caution be exercised regarding interpretation of the results.

Keywords Twin studies · Equal environment assumption · Heritability · Identity by descent

Edited by Elizabeth Prom-Wormley.

✉ David M. Evans
d.evans1@uq.edu.au

[1] The University of Queensland Diamantina Institute, The University of Queensland, Level 7, 37 Kent St, Brisbane, Australia

[2] QIMR Berghofer Medical Research Institute, Brisbane, Australia

[3] School of Biomedical Science, Institute of Health and Biomedical Innovation, Faculty of Health, Queensland University of Technology (QUT), Brisbane, Australia

[4] Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

[5] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

[6] Translational Research Institute, Woolloongabba, QLD 4102, Australia

Published online: 13 February 2021

Introduction

The CTM design relies on the fundamental assumption that trait relevant environmental covariation within monozygotic (MZ) pairs is the same as that found within dizygotic (DZ) pairs-so that zygosity differences in twin covariances must be due to genetic factors and uncontaminated by the environment. This assumption is termed the Equal Environment Assumption (EEA) (Evans and Martin 2000). However, several studies have shown that MZ twins do, in fact, experience more similar environments than DZ twins. As children this may take the form of being dressed alike, sharing the same room, friends and classes (Loehlin and Nichols 1976). Among adults, MZ twins may cohabit more frequently or live nearer together, potentially increasing their sharing of environmental factors. However, where measures of environmental similarity have been made, they rarely associate with pair similarities for outcome traits of interest such as personality or psychiatric disorders (Hettema et al. 1995; Kendler et al. 1993a, b; Loehlin and Nichols 1976). Despite this lack of association between measured treatment similarity and resemblance for other traits, some still claim that the EEA has been violated, and that this may be why heritability estimates from CTM studies appear to be larger than those estimated from genomic data. This discrepancy-termed 'missing heritability' (Maher 2008) is one reason that the EEA is one of the most debated subjects within behavior genetics.

Several early studies set out to test the EEA and concluded, at least for personality, cognitive and psychiatric traits, that the assumption of equal environments seems to stand up well to empirical testing (for reviews see Kendler et al. (1993a, b) and Heath et al. (1989)). Two key observations resulted from early testing of the EEA, the first, that differences in environments between MZ and DZ twins pairs do not appear to bias twin studies and cause inflated heritability estimates unless this greater similarity of experience is etiologically important for the trait under study (Heath et al. 1989; Plomin et al. 1976). Second, the more similar environments that MZ twins experience may in fact be a result of them actively selecting or creating more similar responses from their environment, in which case their increased environmental similarity is a form of gene-environment correlation and not a violation of the EEA (Kendler et al. 1993a, b; Lytton 1977). Indeed the EEA is only violated when MZ twins are passive recipients of more similar environments than DZ twin pairs (Heath et al. 1989). It should also be noted that greater social interaction between MZ pairs than between DZ pairs would (given that some genetic variation is present as well) lead to greater total phenotypic variation in MZ than DZ pairs (Eaves 1976). Substantial effects of this type would lead to poor fit of the usual ACE model, and would usually be detected in preliminary analyses to test whether twins' means and variances differ by zygosity. Again, empirical evidence for such differences is scant; parental ratings of their children's activity levels is one example, though this may be due to parents' rating style differences rather than to sibling interaction (Bartels et al. 2004).

One of the key questions arising from early studies' testing of the EEA was whether increased environmental similarity led to an increased phenotypic concordance between twins or whether excess environmental similarity can be attributed to their genetic identity. For instance, more similar twin pairs may choose to have more frequent social contact, or, on the other hand, more frequent social contact may lead to greater similarity – a question that could be resolved using longitudinal data. Kendler et al. (1993a, b) found that similarity of childhood environment or contact as adults was unrelated to similarity in panic disorder between twins while Posner et al. (1996) showed that attitude similarity leads to contact and that similarity is both genetically and environmentally based. Genetic predisposition to select particular environments, with consequent greater MZ than DZ environmental similarity, does not constitute a violation of the EEA. Rather mechanisms, whereby an individual's (genetically influenced) phenotype impacts their environment, which then feeds back and influences aspects of that same individual's phenotype, underlie the concept of the "Extended Phenotype" as described by Dawkins (1982).

Other studies made use of mistaken zygosity diagnosis to test the EEA. If parents treat MZ twins more similarly than DZ twins based on the preconceived notion that MZ twins are more similar than DZ twins, then trait similarity should be a function of perceived zygosity (Gunderson et al. 2006; Scarr 1968). However, the results from several studies found that misperceived zygosity by twin parents had no significant influence on twin resemblance for several psychiatric disorders (Kendler et al. 1993a, b, 1994). In fact, these identical twins who are thought to be fraternal remain as concordant as MZ twins raised by parents who treated them as identical (Kendler et al. 1993a, b).

Researchers have continued to debate into the 21st century whether or not the equal environment assumption is valid (Conley et al. 2013; Felson 2014; Fosse et al. 2015; Joseph 1998, 2014), leading to several new study designs to directly test the EEA. These included advancements on the stratification approach (where twin similarity is tested as a function of cohabitation or amount of social contact) (Eaves et al. 2003; Eriksson et al. 2006; LoParo and Waldman 2014) and the misclassified twin design (Conley et al. 2013), multivariate data modelling (Derks et al. 2006), the Children of Twins design (Koenig et al. 2010) and exploratory and confirmatory factor analysis (Mitchell et al. 2007). Results across approaches and datasets have generally supported the

validity of the EEA for a range of traits including externalizing disorders, political attitudes, cognitive and psychiatric phenotypes. Some studies reported EEA violations in a small percentage of variables tested. However, the authors concluded that the violations' impact on heritability estimates was small and was unlikely to seriously invalidate their results (Barnes et al. 2014; Felson 2014; Littvay 2012).

While all of the studies cited so far have tested the EEA using environmental similarity between twin pairs after birth, a remaining criticism of the EEA is the assumption that the pre-natal environment of twins is equal. It has been alleged that the circumstances of MZ twin gestation (particularly whether twins are mono- or di-chorionic) may mean that some diseases are more common in MZ than DZ twins (Martin et al. 1997; Phillips 1993). While early studies showed no clear consensus in either direction (Christensen et al. 1995; van den Oord et al. 1995), a 2016 study concluded that the influence of the intrauterine prenatal environment on the MZ twin correlations is small and limited to a few phenotypes (Van Beijsterveldt et al. 2016).

Lastly, recent advances in genotyping technologies and methodologies have given rise to studies that lend weight to the generalizability of estimates from twin studies when they are extended to include non-twin participants. Early on, the 'Virginia 30,000' studies, which included both twins and relatives (parents, siblings, spouses and children) found that the estimates of total genetic and environmental variances were very similar to those estimates solely from twin data (Maes et al. 1997; Truett et al. 1994). Furthermore, in a study using only sibling pairs, Visscher et al (Visscher et al. 2006) conducted an 'assumption free' estimation of heritability using identity-by-descent to estimate phenotypic similarity between sibling pairs. They found that the estimated heritability for height was 0.8, consistent with that estimated by previous twin studies using the CTM. In a genetic study of cardiac conduction traits, Nolte et al. (2017) compared heritability estimates obtained using the classical twin design with those obtained using genetic restricted maximum likelihood (G-REML) analysis (Yang et al. 2011) of the same data (Nolte et al. 2017). For the G-REML analyses, the authors estimated genetic variance components using two genetic relationship matrices- the first consisting of average genome-wide identity by state between all pairwise combinations of individuals in the dataset, and the second containing the same information but setting the relevant matrix element to zero for any pair of individuals who were not closely related (i.e. those pairs of individuals with the genome-wide proportion of alleles identical by state $< 5\%$). Using this method, the combined variance explained by both genetic relatedness matrices divided by the phenotypic variance is equal to the narrow sense heritability of the trait under study and importantly doesn't depend on the EEA (Zaitlen et al. 2013). The authors found that for most traits there was little difference in heritability estimates between the classical twin design and the G-REML analyses, which they interpreted as evidence that the EEA was likely to hold for the majority of these traits.

In the following manuscript we introduce a new method to test the validity of the EEA under strong assumptions with respect to trait etiology and the quality of genetic information available for modeling the genomic similarity between the DZ twin pairs (i.e. no epistatic variance, IBD sharing in DZ twins estimated accurately etc.). Our model uses IBD sharing in DZ twin pairs to estimate heritability and liberate degrees of freedom for estimating separate variance components for MZ and DZ twins. The model builds upon seminal work by Peter Visscher and Nick Martin who were first to estimate heritability in sibling pairs by modelling trait similarity between sibling pairs as a function of genome-wide IBD sharing (Visscher et al. 2006). The new model also has an advantage over many previous methods of testing the EEA in that no specific measures of environmental similarity need to be quantified and correlated with twin similarity, as all environmental factors relevant to the trait under study are tested implicitly in the model. Better still, it quantifies and controls for the effects of possible EEA failures.

Methods

Model Formulation

Our idea is to enhance the CTM by incorporating empirical genome-wide IBD sharing between DZ twins into the model in order to detect possible violations of the EEA. Variation in IBD sharing between DZ twin pairs alone can be used to estimate narrow sense heritability (Visscher et al. 2006). Estimation in this context is achieved by modelling DZ phenotypic pair similarity as a function of genome-wide IBD sharing. The intuition is that, if a trait is genetically influenced, then DZ pairs who share a greater proportion of their genome IBD, should also be more phenotypically similar. Assuming that IBD sharing between randomly selected DZ twin pairs is independent of the environment, estimates of narrow sense heritability derived from DZ pairs alone should be insensitive to violations of the EEA. This contrasts with heritability estimates derived from the CTM where increased environmental similarity in MZ pairs is expected to increase their correlation relative to DZ pairs (and hence mimic genetic effects). Thus, the inclusion of IBD information from DZ twin pairs in the CTM to estimate additive genetic contributions means that information from the twin phenotypic variances, and MZ and DZ intra-pair covariances can now be used to estimate additional parameters, including (under certain assumptions described more in detail below) whether the degree of environmental sharing differs between MZ and DZ twin pairs. We call this modification to the CTM

the "Augmented Classical Twin Design" (Fig. 1, Supplementary Fig. 1).

In this article, we examine the performance of the Augmented Classical Twin Design and its ability to detect violations of the EEA. In Fig. 1, twin phenotypes are modelled as a function of latent additive genetic (A), common environmental (C) and unique environmental (E) factors. Under this model, and according to standard biometrical genetics theory, the covariance between latent additive genetic factors in MZ twins is fixed to V_A, whilst the covariance between latent additive genetic factors in DZ pairs is set to $\hat{\pi}_i \times V_D$, where $\hat{\pi}_i$ is defined as the estimated proportion of alleles shared IBD across the genome for the *i*th DZ pair. $\hat{\pi}_i$ can be estimated empirically using genome-wide microsatellite data or genome-wide SNP data from microarrays in DZ twin pairs with or without parents. $\hat{\pi}_i$ can be estimated using a variety of software programs, but importantly must include contributions from the X chromosome as well as the autosomes (i.e. since X chromosomal loci contribute to heritability their effect must also be modelled) (Manichaikul et al. 2010). The covariance between common environmental factors is fixed to V_{C_MZ} and V_{C_DZ} across MZ and DZ twin pairs respectively, and the correlation between unique environmental factors is fixed to zero. Separate common environmental (V_{C_MZ}, V_{C_DZ}) and unique environmental variance components (V_{E_MZ}, V_{E_DZ}) are estimated in MZ and DZ twins, with the constraint that the total environmental variance across MZ and DZ pairs is equal (i.e. $V_{C\_MZ} + V_{E\_MZ} = V_{C\_DZ} + V_{E\_DZ}$). A possible test of the EEA assumption under this formulation is whether the magnitudes of the common and unique environmental variance components can be equated across MZ and DZ twins (i.e. $V_{C\_MZ} = V_{C\_DZ}$ and $V_{E\_MZ} = V_{E\_DZ}$). Note that this comparison involves a one degree of freedom test, given that the model also constrains the total environmental variance to be equal across MZ and DZ twin pairs.

Before proceeding, we note that an equivalent, although perhaps more intuitively appealing formulation of this model is displayed in the Supplementary Fig. 1. Under this model, phenotypes are modelled as a function of a latent additive

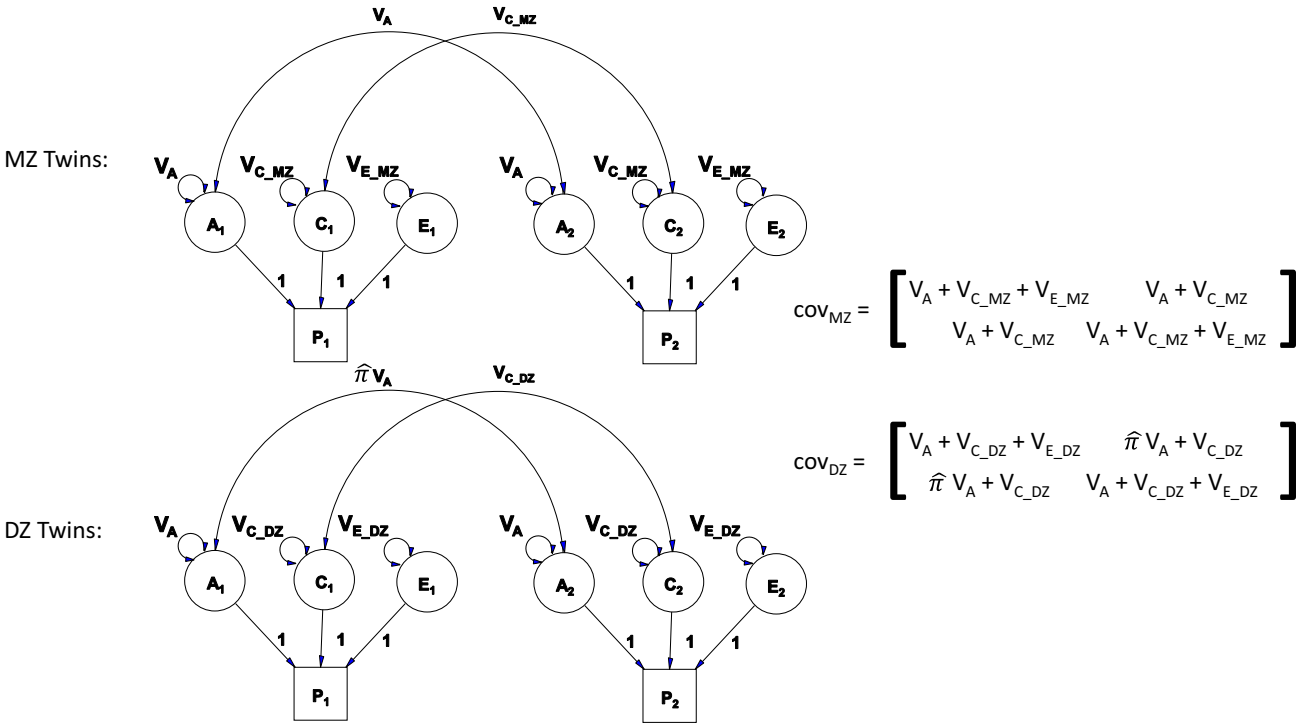


Fig. 1 The Augmented Classical Twin Design. Phenotypes for each twin (P_1, P_2) are modelled as a function of latent additive genetic (A), common environmental (C) and unique environmental (E) sources of variation. The variance of each latent variable is estimated and the path coefficients between the latent and observed variables are constrained to one. The additive genetic covariance between MZ twins is estimated as V_A whilst the additive genetic covariance between the *i*th DZ twin pair is modelled as a function of their genome-wide proportion of alleles IBD ($\hat{\pi}_i$) multiplied by the additive genetic variance. Common environmental and unique environmental variables in MZ and DZ twin pairs are allowed to have different variances with the constraint that the total environmental variance in MZ and DZ twins is the same (i.e. $V_{C\_MZ} + V_{E\_MZ} = V_{C\_DZ} + V_{E\_DZ}$). The expected phenotypic covariances under the full model for each zygosity are displayed. Under this formulation, a test of the EEA is whether common and unique environmental variance components can be constrained equal across zygosities. We note that this model is essentially the same as in Supplementary Fig. 1, but parameterized in a different way. Note that the environmental correlation between MZ twins and between DZ twins can be estimated post-hoc using the formulae: $\hat{r}_{E\_MZ} = \frac{VC\_MZ}{VC\_MZ+VE\_MZ}$ and $\hat{r}_{E\_DZ} = \frac{VC\_DZ}{VC\_DZ+VE\_DZ}$

genetic variable and a single latent environmental variable. The correlation between latent additive genetic factors in MZ twins is fixed to one, whilst the correlation between latent additive genetic factors in DZ pairs is set to $\hat{\pi}_i$,. The variance of the latent environmental factor is estimated (V_E), as well as the correlation between the environmental factors in MZ (r_{E_MZ}) and DZ twin pairs (r_{E_DZ}). If all environmental factors are unshared, then $r_{E\_MZ} = r_{E\_DZ} = 0$. Conversely, in the hypothetical (but impossible) situation that all environmental factors were shared amongst twin pairs, then $r_{E\_MZ} = r_{E\_DZ} = 1$. This parameterization is intuitively appealing in that violations of the EEA should manifest as different environmental correlations between MZ and DZ twins (note that most of the time we would expect excess MZ similarity i.e. $r_{E\_MZ} > r_{E\_DZ}$ although this may not necessarily be the case for all phenotypes e.g. perinatal phenotypes where unequal blastocyst sharing may contribute to differences in MZ twin pairs (Martin et al. 1997). Conversely, when the EEA is satisfied, then we expect that $r_{E\_MZ} = r_{E\_DZ}$. It follows then, that a possible test of the validity of the EEA is whether the estimated correlation between latent environmental factors is the same in MZ and DZ twins.

We note, that although these alternative parameterizations may look different on the surface, they are in fact equivalent formulations of the same underlying model. Indeed, the correlation between latent environmental factors in MZ and DZ twin pairs can be estimated post-hoc in the model displayed in Fig. 1 using the following equations which relate these quantities to the estimated variance components:

$$\hat{r}_{E\_MZ} = \frac{\hat{V}_{C\_MZ}}{\hat{V}_{C\_MZ} + \hat{V}_{E\_MZ}} \quad (1)$$

$$\hat{r}_{E\_DZ} = \frac{\hat{V}_{C\_DZ}}{\hat{V}_{C\_DZ} + \hat{V}_{E\_DZ}} \quad (2)$$

However, in this manuscript we only present results for the model displayed in Fig. 1. This is because we found it to have superior estimation properties to the alternative formulation (Supplementary Fig. 1) in our simulated data. Briefly, the model in Fig. 1 permits negative variance components whereas the other model forces variances to be positive which can result in estimation difficulties in some situations.

Finally, we build intuition for why parameters in the Augmented Classical Twin Design are identified. Information on estimating the additive genetic variance component predominantly comes from differences in DZ pair similarity as a function of their genome-wide IBD sharing. Given that V_A is identified, it is then possible to uniquely resolve V_{C_DZ} and V_{E_DZ} from the DZ variance and covariance respectively. In MZ twins, the difference between the MZ covariance and trait variance is sufficient to identify V_{E_MZ}. Given that V_A is already identified, it follows that V_{C_MZ} can be identified from the difference between the MZ covariance and V_A. Thus, it follows that all parameters are identified, and the model as a whole is identified.

Simulations Investigating Power and Type I Error

In order to investigate the properties of the Augmented Classical Twin Design we conducted a series of simulations. Phenotypic values for the *i*th twin pair (P_{1i}, P_{2i}) were generated as a function of additive genetic and environmental variables:

$P_{1i} = a \times A_{1i} + e \times E_{1i}$.

$P_{2i} = a \times A_{2i} + e \times E_{2i}$.

Additive genetic variables (A_1, A_2) were drawn from a bivariate normal distribution with mean zero and variance one in which the correlation was set to one in the case of MZ twin pairs and $\hat{\pi}_i$ for the *i*th DZ twin pair (see below). Environmental variables (E_1, E_2) were drawn from a second, independent, bivariate normal distribution with mean zero and variance one where the correlation was set to r_{E_MZ} in the case of MZ twin pairs and r_{E_DZ} in the case of DZ twin pairs. The genome-wide proportion of alleles shared IBD in DZ twin pairs was modelled by sampling from a random normal distribution with mean = 0.5 and standard deviation = 0.039. These parameter values represent asymptotic values derived from quantitative genetics theory, although empirical estimates of these quantities obtained from real studies involving genome-wide panels of microsatellite markers show remarkable concordance with the theoretical expectations (Visscher et al. 2006). Given that the information content of dense genome-wide SNP microarrays is greater than traditional microsatellite panels (John et al. 2004), we would expect that $\hat{\pi}$ estimates derived from genome-wide SNP data would yield similar distributions. Trait heritability was varied across simulations by modifying the value of the path coefficients *a* and *e* so that the additive genetic variance ($V_A = a^2$), and environmental variance ($V_E = e^2$) also changed with the constraint that $V_A + V_E = 1$.

We initially investigated four combinations of sample size (20,000 MZ and 20,000 DZ pairs; 20,000 MZ and 50,000 DZ pairs; 50,000 MZ and 20,000 DZ pairs; 50,000 MZ and 50,000 DZ pairs) and three conditions corresponding to traits with low, moderate or high heritability ($V_A = 0.2$, $V_E = 0.8$; $V_A = 0.5$, $V_E = 0.5$; $V_A = 0.8$, $V_E = 0.2$). Within each of these twelve conditions, we varied the variance explained by common and unique environmental sources of variation in MZ and DZ twin pairs, with the constraint that V_{C_MZ} was always $\geq V_{C\_DZ}$. Following our initial simulations which suggested that results might be relatively insensitive to the number of MZ twin pairs in the analysis, we investigated

three additional conditions related to sample size (i.e. 20,000 DZ pairs along with either 100 MZ pairs, 1,000 MZ pairs or 10,000 MZ pairs).

We fitted the full model to the simulated data. In all cases variance components were estimated without constraining their values to be positive. We also fitted a nested sub-model where the environmental variance components were constrained equal between MZ and DZ twins (i.e. $V_{C\_MZ} = V_{C\_DZ}$ and $V_{E\_MZ} = V_{E\_DZ}$). The difference in fit between full and nested sub-models was evaluated against a chi-square distribution with one degree of freedom. We performed 1,000 replicates for each condition and calculated mean parameter estimates, standard errors, and Type I error rates/power across the replicates. All simulations were conducted in R statistical software using the software package OpenMx (Neale et al. 2016). An example script detailing our simulations is included in the Supplementary Materials.

Extension to Include Genetic Dominance

Under our framework, information on potential violations of the EEA is obtained from the difference between the estimate of the additive genetic variance obtained by modelling IBD sharing between DZ twin pairs and that obtained from the MZ and DZ intra-pair correlations. Increased similarity of MZ twins due to EEA violations are expected to result in higher heritability estimates from the CTM than those from studies using IBD sharing in DZ twin pairs. Any process that makes MZ twins more similar to each other than DZ twins, would be expected to produce evidence for violation of the EEA. Such processes would include (but are not limited to) genetic non-additivity due to dominance and/or epistasis. We were therefore interested in how the presence of genetic non-additivity might adversely affect the Augmented Classical Twin Design, and whether we could assuage its influence by modelling it appropriately in our framework.

We therefore included a genetic dominance variance component (Fig. 2). Our parameterization is exactly the same as previously, except for the additional dominance variance component. Under this formulation, the covariance between dominance genetic factors is set to V_D in MZ twins, and $\widehat{p}_{2,i} \times V_D$ in DZ twin pairs, where $\widehat{p}_{2,i}$ is the estimated coefficient of dominance variance, defined as the estimated proportion of the genome in which the ith DZ pair shares both alleles IBD. It has been shown previously that the power to detect genetic dominance is low using methods that involve IBD sharing between sibling pairs, in part because $\widehat{\pi}_i$ and $\widehat{p}_{2,i}$ are highly correlated, and therefore so are estimates of V_A and V_D (Dominicus et al. 2006; Visscher et al. 2006). However, under our model, even though the power to resolve V_A and V_D may be low, it is possible that the total genetic variance (i.e. $V_A + V_D$) may be estimated relatively precisely, in which case there may still be power to reject violations of the EEA.

In order to investigate the effect of genetic dominance on the Augmented Classical Twin Design, we simulated 20,000 MZ and 20,000 DZ twin pairs. Trait values were simulated as previously described, except for the additional variance component due to genetic dominance (D):

$P1_i = a \times A_{1i} + d \times D_{1i} + e \times E_{1i}$.

$P2_i = a \times A_{2i} + d \times D_{2i} + e \times E_{2i}$.

We simulated conditions where additive genetic factors were responsible for 40 % of the trait variance ($V_A = a^2 = 0.4$), dominance genetic factors were responsible for 10 % of the trait variance ($V_D = d^2 = 0.1$), and common and/or environmental factors the remaining 50 % of the variance. For each DZ twin pair, the genome-wide proportion of alleles IBD and the genome-wide coefficient of dominance were sampled from a bivariate normal distribution with $\mu = \begin{bmatrix} 0.5 \\ 0.25 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{matrix} 0.039^2 & 0.00153 \\ 0.00153 & 0.044^2 \end{matrix}$ (correlation between $\widehat{\pi}$ and $\widehat{p} \cong 0.89$). These quantities are based on theoretical estimates of IBD sharing between siblings (Visscher et al. 2006). We analysed fifteen different combinations of common and unique environmental variances in the MZ and DZ twins subject to the constraint that V_{C_MZ} was always $\geq V_{C\_DZ}$).

Results

Bias and Type I Error Rates

Mean parameter estimates, Type I error rates, and statistical power for the simulated conditions that don't involve genetic dominance are displayed in Supplementary Table 1. The quantities V_A, V_{C_MZ}, V_{C_DZ}, V_{E_MZ} and V_{E_DZ} were estimated with little bias when the full model was fitted to the data. This implies that even in situations where the EEA is violated, the Augmented Classical Twin Design returns asymptotically unbiased estimates of trait heritability and variance components, provided that strong assumptions including no genotype-environment covariance, no assortative mating, no GxE interaction, no epistasis, and bivariate normality have not been violated. Nevertheless, there was still substantial sampling variability in the estimated variance components across sample replicates (i.e. V_A, V_{C_MZ}, V_{C_DZ}, and V_{E_DZ}, although not V_{E_MZ}), consistent with the low power to estimate heritability using IBD information in sibling pairs alone (Visscher et al. 2006), and highlighting the fact that very large numbers of DZ twins are required in order to produce precise variance component estimates.

We also examined post-hoc estimates of r_{E_MZ} and r_{E_DZ} calculated using estimates of the variance components via

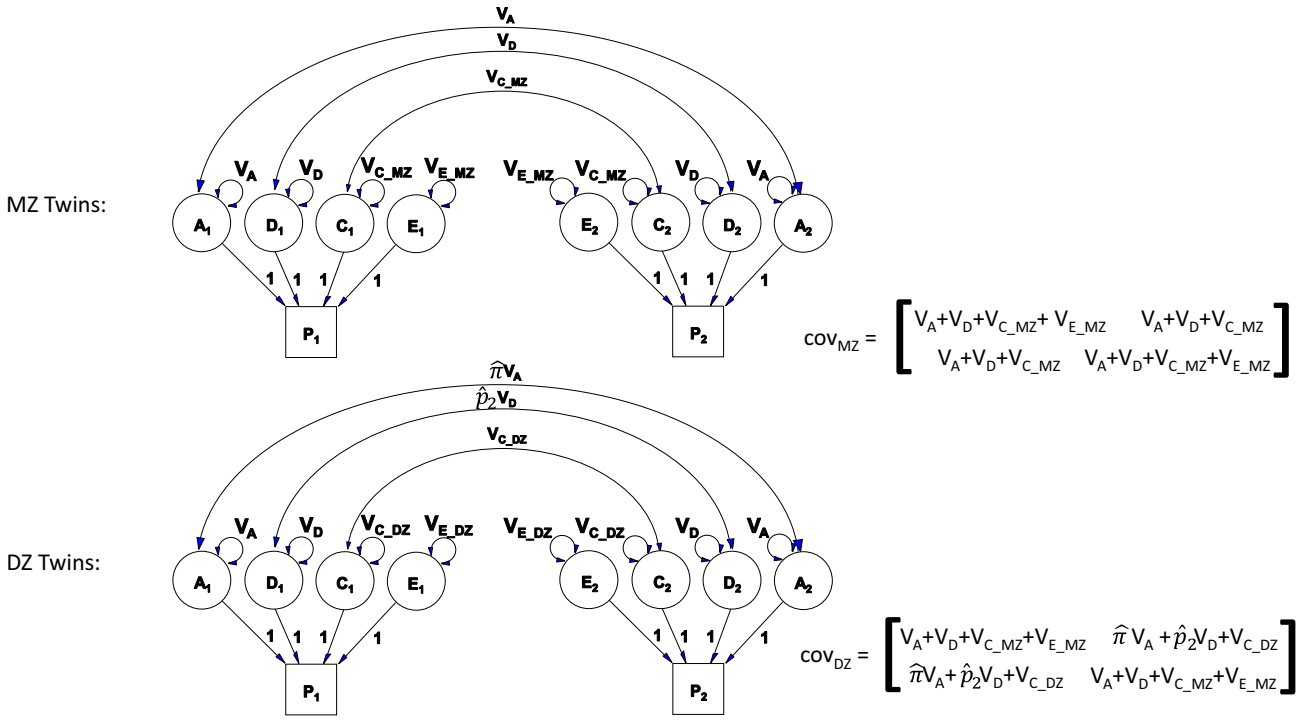


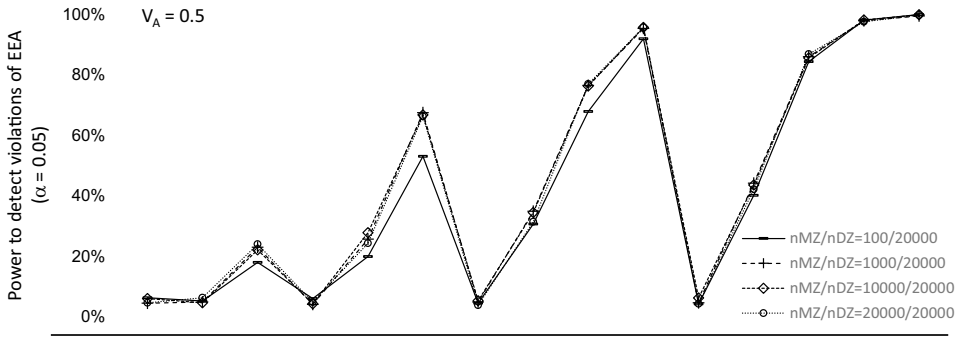
Fig. 2 Extension of the Augmented Classical Twin Design to Include Genetic Dominance. Phenotypes for each twin (P_1, P_2) are modelled as a function of latent additive genetic (A), dominance genetic (D), common environmental (C) and unique environmental (E) sources of variation. The variance of each latent variable is estimated and the path coefficients between the latent and observed variables are constrained to one. The additive genetic covariance between MZ twins is estimated as V_A whilst the additive genetic covariance between the *i*th DZ twin pair is modelled as a function of their estimated genome-wide proportion of alleles IBD ($\hat{\pi}_i$) multiplied by the additive genetic variance. The dominance genetic covariance between MZ twins is estimated as V_D whilst the dominance genetic covariance between the *i*th DZ twin pair is modelled as a function of the estimated proportion of the genome in which they share both alleles IBD ($\hat{p}_{2,i}$) multiplied by the dominance genetic variance. Common environmental and unique environmental variables in MZ and DZ twin pairs are allowed to have different variances with the constraint that the total environmental variance in MZ and DZ twins is the same (i.e. $V_{C\_MZ} + V_{E\_MZ} = V_{C\_DZ} + V_{E\_DZ}$). The expected phenotypic covariances under the full model for each zygosity are displayed. Under this formulation, a test of the EEA is whether common and unique environmental variance components can be constrained equal across zygosities

Eqs. 1 and 2. In general, estimates of r_{E_MZ} and r_{E_DZ} were downward biased (r_{E_MZ} often more so). The magnitude of the bias was often less for simulated conditions where V_A was low, V_{C_MZ} and V_{C_DZ} were high, and the number of DZ twin pairs was high. Indeed for many of the conditions where the simulated heritability was high ($V_A = 0.8$) and for a few where the simulated heritability was moderate ($V_A = 0.5$), estimates of r_{E_MZ} and r_{E_DZ} were unstable. These conditions involved replicates where patterns of IBD sharing amongst the DZ twins implied very high estimates of the additive genetic variance (i.e. close to one). This resulted in negative estimates of V_{C_MZ} (or V_{C_DZ}) in order to be compatible with the observed phenotypic covariance between MZ (or DZ) twin pairs. When the estimated unique environmental component was of similar magnitude (but opposite direction), the denominator in the formula for r_{E_MZ} and r_{E_DZ} (see Eqs. 1 and 2) was close to zero, resulting in highly unstable estimates. Replicates exhibiting these features were rare for simulations where heritability was moderate ($V_A = 0.5$; <1 % of replicates) but more common when heritability was high ($V_A = 0.8$; up to ~ 10 % of replicates depending on the condition) in some cases badly skewing average values of the correlation estimates across simulations. Increasing the number of DZ pairs in the analysis by an order of magnitude mitigated this problem (i.e. because the additive genetic variance could be estimated more precisely and so estimates of the common environmental variance components were less likely to be negative- data not shown). Importantly though, Type I error rates for the test of the EEA were appropriate across all conditions, including those where post-hoc estimates of r_{E_MZ} and r_{E_DZ} were biased (Supplementary Tables 1 and 2).

Statistical Power

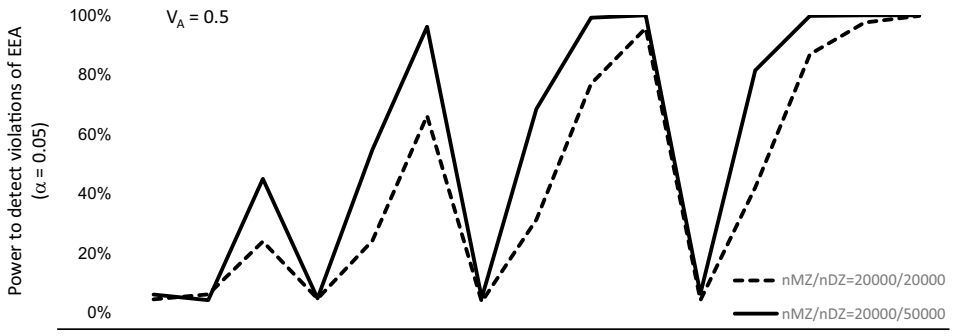
Interestingly, the power to detect violations of the EEA was relatively insensitive to the number of MZ twin pairs in the analysis for the range of conditions that we examined. This is highlighted in Fig. 3, which shows that even relatively few MZ twins (e.g. 100 MZ pairs) produced almost the same

power as much larger numbers. The number of DZ twin pairs is the more critical consideration. Figure 4 illustrates the effect of modifying the environmental similarity in MZ and DZ twins on Type I error rates and statistical power. To assist in interpretation, we present these results showing the two equivalent ways of thinking about environmental sharing (i.e. showing V_{C_MZ} and V_{C_DZ} or alternatively and equivalently, showing r_{E_MZ} and r_{E_DZ}). Unsurprisingly power to detect violations of the EEA increases with increasing number of DZ twin pairs, increasing difference between r_{E_MZ} and r_{E_DZ}, and increasing proportion of the variance due to V_C.



V_{C_MZ}	0	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4
V_{C_DZ}	0	0.1	0	0.2	0.1	0	0.3	0.2	0.1	0	0.4	0.3	0.2	0.1	0
r_{E_MZ}	0	0.2	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.6	0.8	0.8	0.8	0.8	0.8
r_{E_DZ}	0	0.2	0	0.4	0.2	0	0.6	0.4	0.2	0	0.8	0.6	0.4	0.2	0

Fig. 3 Effect of modifying the environmental similarity between MZ and DZ twins on Type I error rates and statistical power for different numbers of MZ twins. We simulated phenotypic values for twin pairs as a function of additive genetic (V_A) and environmental (V_E) variables with varying environmental similarities for varying numbers of MZ pairs and a fixed number of 20,000 DZ pairs (1,000 rounds of simulation for each condition). Power to detect violations of the equal environment assumption (EEA) was estimated by comparing the model fits between the constrained model ($V_{C\_MZ} = V_{C\_DZ}$) and the unconstrained model and counting the proportion of replicates where this achieved significance ($p < 0.05$). To assist in interpretability of the results we label the x-axis using both the proportion of shared environmental variance for MZ (V_{C_MZ}) and DZ twins (V_{C_DZ}), and also (equivalently), the environmental correlations between MZ and DZ twins ($\hat{r}_{E\_MZ}$, $\hat{r}_{E\_DZ}$)



V_{C_MZ}	0	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4
V_{C_DZ}	0	0.1	0	0.2	0.1	0	0.3	0.2	0.1	0	0.4	0.3	0.2	0.1	0
r_{E_MZ}	0	0.2	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.6	0.8	0.8	0.8	0.8	0.8
r_{E_DZ}	0	0.2	0	0.4	0.2	0	0.6	0.4	0.2	0	0.8	0.6	0.4	0.2	0

Fig. 4 Effect of modifying the environmental similarity between MZ and DZ twins on Type I error rates and statistical power. We simulated phenotypic values for twin pairs as a function of additive genetic (V_A) and environmental (V_E) variables with varying environmental similarities for four combinations of sample size (1000 rounds of simulation for each condition). Power to detect violations of the equal environment assumption (EEA) was estimated by comparing the model fits between the constrained model ($V_{C\_MZ} = V_{C\_DZ}$) and the unconstrained model and counting the proportion of replicates where this achieved significance ($p < 0.05$). To assist in interpretability of the results we label the x-axis using both the proportion of shared environmental variance for MZ (V_{C_MZ}) and DZ twins (V_{C_DZ}), and also (equivalently), the posthoc environmental correlations between MZ and DZ twins ($\hat{r}_{E\_MZ}$, $\hat{r}_{E\_DZ}$)

In interpreting these simulations, it is useful to consider how they relate to the proportion of variance explained by genetic factors that would have been estimated had an ordinary ACE/ADE model been fitted to the data instead (i.e. and would therefore contribute to missing heritability if the ordinary ACE/ADE model was used incorrectly to estimate trait heritability). We can examine this by subtracting the proportion of genetic variance expected under a reduced ACE (or ADE) model from the true simulated values. Let V_P represent the phenotypic variance, COV_{MZ} the phenotypic covariance between MZ pairs, and COV_{DZ} the phenotypic covariance between DZ pairs. Under the reduced model, and when the $COV_{MZ} <= 2 \times COV_{DZ}$ (as would be expected for a trait with A C and E variance components), the estimated additive genetic variance has expectation:

$E(\widehat{V}_A) = 2COV_{MZ} - 2COV_{DZ} = 2(V_A + V_{C\_MZ}) - 2(½V_A + V_{C\_DZ}) = V_A + 2V_{C\_MZ} - 2V_{C\_DZ}$.

The difference in the proportion of variance explained by genetic factors can then be calculated as:

$$\frac{E\left(\widehat{V}_A\right) - V_A}{V_P} = \frac{2V_{C\_MZ} - 2V_{C\_DZ}}{V_P}$$

When the $COV_{MZ} > 2 \times COV_{DZ}$ (consistent with an ADE model, as long as $COV_{MZ} <= 4 \times COV_{DZ}$):

$E(\widehat{V}_A) = 4COV_{DZ} - COV_{MZ} = 4(½V_A + V_{C\_DZ}) - (V_A + V_{C\_MZ}) = V_A + 4V_{C\_DZ} - V_{C\_MZ}$.

and:

$E(\widehat{V}_D) = 2(COV_{MZ} - 2COV_{DZ}) = 2[V_A + V_{C\_MZ} - 2(½V_A + V_{C\_DZ})] = 2V_{C\_MZ} - 4V_{C\_DZ}$.

The difference in the proportion of variance explained by genetic factors can then be calculated as:

$$\frac{E\left(\widehat{V}_A\right) + E(\widehat{V}_D) - V_A}{V_P}$$

We calculated this quantity for all the models we have run and included it in the last columns of Supplementary Tables 1 and 2. Figure 5 shows power to detect violations of the EEA as a function of this difference for a representative selection of the conditions we have examined. Figure 5 shows that very large numbers of DZ pairs (i.e. > 50,000) will be required to detect violations of the EEA assumption that explain low absolute differences in the proportion of genetic variance between the models (i.e. 10 % of the missing heritability) with moderate power.

The Effect of Genetic Dominance

As predicted, the inclusion of genetic dominance increased MZ intra-pair similarity relative to DZ intra-pair similarity and resulted in increased type I error, and biased estimates for variance components V_A, V_{C_DZ} and V_{E_DZ} (V_{C_MZ} and V_{E_MZ} were unbiased in the simulations we examined) and post-hoc estimates of the MZ and DZ environmental correlations (Supplementary Table 3). Specifically, average estimates of V_A were inflated relative to its expected value by

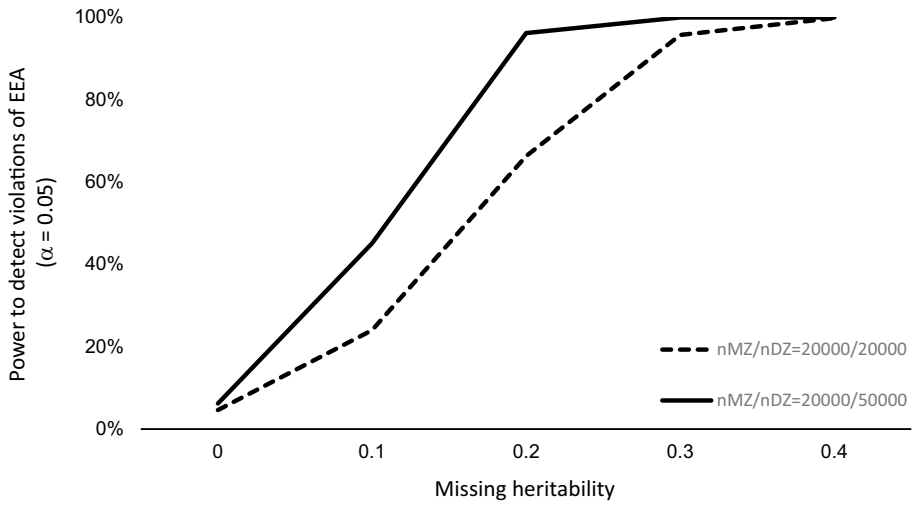


Fig. 5 Power to detect violations of the equal environment assumption as a function of missing heritability. Missing heritability is defined as the difference between the additive genetic variance estimated in the Classic Twin Model and that estimated in the Augmented Classic Twin Design. In this figure, missing heritabilities were calculated from simulations of phenotypic values for twin pairs with additive genetic variance (V_A) of 0.5, varying environmental correlation between MZ pairs (r_{E_MZ} = 0, 0.2, 0.4, 0.6, and 0.8) and a fixed environmental correlation of 0 between DZ pairs (r_{E_DZ} = 0), resulting in respective missing heritabilities of 0, 0.1, 0.2, 0.3, and 0.4. Power was estimated in the simulations using two combinations of sample size

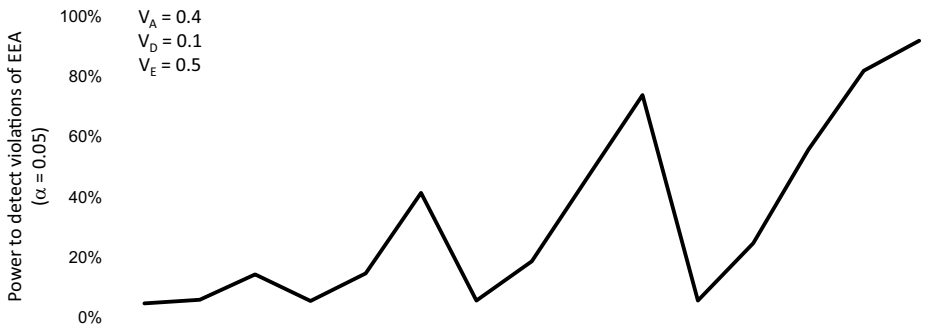
a quantity equal to the amount of genetic dominance simulated, V_{C_DZ} was underestimated, and V_{E_DZ} overestimated. Interestingly, fitting a dominance term in the model was sufficient to account for bias in estimates of variance components under the conditions that we examined, although power to detect violations of the EEA was reduced relative to similar situations where dominance did not affect the trait (Fig. 6).

Discussion

We have shown that it is possible using genome-wide IBD sharing to model the environmental correlation between MZ and DZ twin pairs and subsequently conduct valid tests of the EEA assumption using an Augmented Classical Twin Design (provided certain strong assumptions regarding trait etiology are met - see below). Our model extends seminal work by Visscher and colleagues who were the first to use empirical genome-wide IBD sharing in sibling pairs to estimate heritability (Visscher et al. 2006). Our contribution is to realize that the addition of genome-wide IBD information to the CTM allows investigators to estimate different environmental components of variance for MZ and DZ twins and use this information to formally test the EEA.

Our simulations show that the power of the Augmented Classical Twin Design to detect violations of the EEA depends mostly on the number of DZ pairs in the analysis, and is largely insensitive to the number of MZ twin pairs. To understand this result intuitively, note first that information on V_{E_MZ} comes from the difference between the overall trait variance and the MZ covariance. It is well known that there is high power to estimate this component, even when there is only a very small number of MZ twin pairs (Martin et al. 1978). In contrast, power to resolve V_A comes primarily from the correlation between IBD sharing and trait similarity in DZ twin pairs. Therefore, the most important factor in estimating V_A is the total number of DZ twin pairs. When V_A and V_{E_MZ} are estimated well, it implies that V_{C_MZ} will also tend to be estimated well. Likewise, large numbers of DZ pairs are also essential to ensure that V_{C_DZ} and V_{E_DZ} are estimated with precision (Visscher et al. 2006). In other words, except for V_{E_MZ} (which requires only a relatively small number of MZ twin pairs to be estimated precisely), the other parameters require large numbers of DZ pairs to be estimated well. The corollary is that the EEA can be tested using the Augmented Classical Twin Design using relatively few MZ twin pairs and many genotyped DZ twin pairs.

Violation of the EEA, in the direction of increased MZ similarity, is expected to lead to inflated estimates of heritability, and may be one of the explanations for the "missing heritability" phenomenon (Maher 2008; Manolio et al. 2009). The conditions that we have examined via simulation translate to relatively large differences between the heritability estimated from the CTM and the true heritability (Fig. 5). Indeed, we have shown that in many instances a large sample consisting of 50,000 DZ pairs provides only moderate power to detect relatively large violations of the EEA and missing



V_{C_MZ}	0	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4
V_{C_DZ}	0	0.1	0	0.2	0.1	0	0.3	0.2	0.1	0	0.4	0.3	0.2	0.1	0
r_{E_MZ}	0	0.2	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.6	0.8	0.8	0.8	0.8	0.8
r_{E_DZ}	0	0.2	0	0.4	0.2	0	0.6	0.4	0.2	0	0.8	0.6	0.4	0.2	0

Fig. 6 Effect of modifying the size of the MZ and DZ environmental correlations on Type I error rates and statistical power to detect violations of the equal environment assumption when genetic dominance contributes to trait variability. We simulated phenotypic values for twin pairs as a function of additive genetic (V_A), dominance genetic (V_D), and environmental (V_E) variables with varying environmental correlation between MZ pairs (r_{E_MZ}) and between DZ pairs (r_{E_DZ}) for 20,000 MZ and 20,000 DZ pairs and each 1,000 rounds of simulation. Power to detect violations of the equal environment assumption (EEA) were estimated by comparing the model fits between the constrained model ($V_{C\_MZ} = V_{C\_DZ}$) and the unconstrained model and counting the proportion of replicates where this achieved significance ($p < 0.05$). To assist in interpretability of the results we label the x-axis using both the proportion of shared environmental variance for MZ (V_{C_MZ}) and DZ twins (V_{C_DZ}), and also (equivalently), the environmental correlations between MZ and DZ twins ($\hat{r}_{E\_MZ}$, $\hat{r}_{E\_DZ}$)

heritability. It is unclear how common such extreme violations of the EEA might be in reality. For example, in the case of height, there is little deviation between heritability as estimated using the CTM and heritability as estimated from IBD sharing between sibling pairs (Hemani et al. 2013), although it might be expected *a priori* that a trait like adult height is unlikely to be greatly affected by violations of the EEA, or by genotype by age interactions since it varies little once adulthood is achieved. For other traits like BMI, the discrepancy between estimates of heritability using these different designs is greater, although such large differences may also be a consequence of age dependent gene expression, decreasing the concordance between sibling pairs. For psychological traits and those primarily of interest to behavior geneticists, the magnitude of violation of the EEA may be greater, although evidence for substantial violations of the EEA using a range of alternative methods is sparse (Evans and Martin 2000; Kendler et al. 1993a, b). The implication, however, is that the power to detect anything other than large violations of the EEA using the Augmented Classical Twin Design is likely to be limited given the current size of twin registries around the world.

We have also shown that it is possible to perform a valid test of the EEA whilst allowing for potential confounding through genetic dominance at the cost of a relatively small decrement to statistical power. This contrasts with the well-known low power of IBD-based tests of genetic dominance, which suffer because of the high correlation between the additive and dominance components of variance (Visscher et al. 2006). The difference is because, although we have low power to resolve genetic components of variance into additive or dominance sources of variation, there is still power to estimate the total variance due to genetic sources of variation, which is the more important consideration when the focus is on estimating the variance due to the environment and the MZ and DZ environmental correlations.

Whilst we were able to examine the effect of genetic dominance on the Augmented Classical Twin Design and subsequently correct for its undesirable effects by including the estimated genome-wide coefficient of dominance variance ($\hat{p}_2$) in our model for DZ twin pairs, we did not examine the effects of genetic epistasis on our model. We expect that failure to model genetic epistasis would have effects similar to those of failure to model genetic dominance on our results. Epistasis-interactions between different genomic loci-would increase similarity amongst MZ twin pairs relative to DZ twin pairs, inducing spurious evidence for violation of the EEA. However, epistatic variance is much more difficult to model in terms of IBD sharing between DZ twin pairs than genetic dominance. There is also no practical limit to the number of loci involved in the epistatic interaction, nor the nature of it (e.g. a two locus interaction includes possible additive × additive, additive × dominant, dominant × additive, and dominant × dominant terms, a three locus interaction even more possibilities, and so on and so forth). It remains to be seen how common epistatic variance is in human populations, however, to date there is only limited empirical evidence that it makes a substantial contribution to variation in complex human traits and diseases (Evans et al. 2011; Genetic Analysis of Psoriasis et al. 2010; Heath et al. 1984). Additionally, estimates of variance associated with higher order interaction terms may correlate substantially with those for additive and/or dominance sharing, so it is possible that additive and dominance variance components already capture some of this additional unmodelled variation.

Regardless, it is critically important that the reader appreciates that our design relies on strong assumptions for its validity. Specifically, *any* unmodelled factor that increases MZ relative to DZ similarity will increase estimates of r_{E_MZ} relative to r_{E_DZ}, and result in "evidence" that the EEA has been violated, regardless of whether this is true in reality. Factors that increase MZ relative to DZ similarity include epistasis as discussed above, but also imperfections in the quality of genetic information used for modelling the genomic similarity between the DZ twin pairs. Measurement error in IBD calculations between DZ pairs will result in underestimation of the additive genetic variance in the DZ twin pairs and could also result in spurious evidence for violation of the EEA. We therefore counsel extreme caution when applying and interpreting the results of the Augmented Classical Twin Design, as significant reductions in model fit may reflect factors other than violations of the EEA if the strong assumptions underlying the model are not satisfied in reality.

It is interesting to speculate whether the power of the Augmented Classical Twin Design could be further improved by including non-twin sibling pairs in the analysis (and estimating separate common environmental and unique environmental variance components for this type of relationship, with the constraint that the total environmental variance is the same as in the twin pairs). We expect that the inclusion of non-twin siblings would assist in more precisely estimating the additive genetic and dominance genetic variance components and consequently would increase power to detect differences in the degree of common environmental sharing between MZ and DZ twin pairs. This consideration is important because we have shown that power to detect violations of EEA is low unless tens of thousands of genotyped DZ pairs are available for analysis. However, twin cohorts/consortia of this size do not exist currently, whereas there are already very large samples of genome-wide genotyped sibling pairs that are available e.g. as part of the Within Families Consortium (Brumpton et al. 2020). The inclusion of sibling pairs in the Augmented Classical Twin Design is likely appropriate for traits that exhibit marked

stability over time, such as adult height. However, caution is required for traits that exhibit more marked temporal variation. Longitudinal twin studies show that many traits exhibit age-dependent expression of genes (Boomsma and Molenaar 1987). In this situation, the inclusion of non-twin siblings (who by definition are of different ages) may artificially lower estimates of the additive genetic variance, and hence produce spurious evidence for violations of the EEA using this model. Indeed age-dependent gene expression may be one reason why heritability estimates derived using IBD sharing in sibling pairs are lower than those obtained from the CTM for traits such as body mass index (Hemani et al. 2013). Nevertheless, it may be possible to model (and test for the existence of) some varieties of gene × age interaction by e.g. including age as a moderator variable in the model (Purcell 2002). We also note that even in the absence of non-twin siblings, it is important that the samples of MZ and DZ pairs in the Augmented Classical Twin Design are matched in terms of age to deal with similar concerns regarding the presence of gene × age interaction.

One of the criticisms levelled at methods designed to detect violations of the EEA is that many do not distinguish between whether twins' more similar treatment is a consequence of their actively eliciting (or locating) more similar responses from their environment, or whether they are just passive recipients of more similar environments. In the former case, the increased environmental similarity that MZ twins experience is not a violation of the EEA, but rather a form of gene-environment covariance, since the environment that twins find themselves in is a function of their genotype. Since environmental similarity due to gene-environmental correlation is a function of IBD sharing in DZ twins, this excess similarity should also contribute to estimates of the additive genetic variance in the Augmented Classical Twin Design. In contrast, if environmental similarity is a passive process and not a consequence of twins' genetic similarity, then environmental similarity will not be correlated with IBD sharing between DZ twin pairs, and will be reflected by different values of the MZ and DZ estimated environmental variance components. In other words, the Augmented Classical Twin Design should be able to distinguish between active and passive genotype-environment covariance.

Another important consideration is whether to include opposite sex twin pairs in the Augmented Classical Twin Design. If the genetic architecture of the trait under examination differs between males and females, then the inclusion of opposite sex DZ pairs will lower estimates of heritability and in turn provide spurious evidence for violation of the EEA. For this reason, we suggest that only same-sex DZ pairs be used in the Augmented Classical Twin Design, unless there is compelling evidence that the genetic etiology of the trait in males and females is the same. Likewise, it is important that users of the model include X chromosome sharing in genome-wide IBD calculations. We make the additional assumption that genetic polymorphisms on the Y chromosome and mitochondria do not make substantial contributions to trait variance.

The Augmented Classical Twin Design could be extended in several ways, most obviously to the multivariate case. Violations of the EEA potentially bias estimates of both genetic variance and the genetic covariance between traits, and consequently the estimated source and structure of the genetic and environmental covariance in structural equation models of twin pairs. Modelling different intra-pair environmental correlations (both within and across traits) for MZ and DZ twin pairs would be one way of estimating and controlling for bias in the analysis of covariance between traits. Another possibility would be to extend the model to binary and ordinal data. It would be a relatively simple matter to model affection status using thresholds and a normal underlying distribution of liability. We predict, however, that many more genotyped DZ twin pairs (perhaps 10 × or more, (Neale and Kendler 1995)) will be required to achieve levels of power similar to those for quantitative traits, and so practical application of this model may not be realistic in the near future.

In conclusion, we have presented an extension to the CTM called the Augmented Classical Twin Design which uses IBD sharing between DZ twin pairs to estimate different common and unique environmental variance components for MZ and DZ twins. It provides a statistical test of whether these variance components can be equated across zygosities and information on whether the EEA assumption holds for the trait of interest. We show through simulation that this model provides unbiased estimates of variance components and valid tests of the EEA under strong assumptions relating to trait etiology which may not be satisfied in real data.

Supplementary Information The online version of this article (https://doi.org/10.1007/s10519-021-10044-0) contains supplementary material, which is available to authorized users.

Acknowledgements D.M.E. and S.E.M are funded by Australian National Health and Medical Research Council Senior Research Fellowships (APP1137714 and APP 1103623). B.L.M is grateful for support from Queensland University of Technology through a QUT Postgraduate Research Scholarship.

Funding D.M.E. and S.E.M are funded by Australian National Health and Medical Research Council Senior Research Fellowships (APP1137714 and APP 1103623).

Compliance with Ethical Standards

Conflict of interest Liang-Dar Hwang, Brittany L. Mitchell, Sarah E. Medland, Nicholas G. Martin, Michael C. Neale andDavid M. Evans report no conflicts of interest.

Research Involving Human and/or Animal Participants This article does not contain any studies with human participants or animals performed by any of the authors.

References

Barnes J, Wright JP, Boutwell BB, Schwartz JA, Connolly EJ, Nedelec JL, Beaver KM (2014) Demonstrating the validity of twin research in criminology. Criminology 52(4):588–626

Bartels M, Boomsma DI, Hudziak JJ, Rietveld MJ, van Beijsterveldt TC, van den Oord EJ (2004) Disentangling genetic, environmental, and rater effects on internalizing and externalizing problem behavior in 10-year-old twins. Twin Res 7(2):162–175

Boomsma DI, Molenaar PC (1987) The genetic analysis of repeated measures. I. Simplex models. Behav Genet 17(2):111–123

Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie GA, Cho Y, Howe LD, Hughes A, Boomsma DI, Havdahl A, Hopper J, Neale M, Nivard MG, Pedersen NL, Reynolds CA, Tucker-Drob EM, Grotzinger A, Howe L, Morris T, Li S, Within-family C, andMe Research T, Auton A, Windmeijer F, Chen WM, Bjorngaard JH, Hveem K, Willer C, Evans DM, Kaprio J, Davey Smith G, Asvold BO, Hemani G, Davies NM (2020) Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. Nat Commun 11(1):3519

Christensen K, Vaupel JW, Holm NV, Yashin AI (1995) Mortality among twins after age 6: fetal origins hypothesis versus twin method. BMJ 310(6977):432–436

Conley D, Rauscher E, Dawes C, Magnusson PK, Siegal ML (2013) Heritability and the equal environments assumption: evidence from multiple samples of misclassified twins. Behav Genet 43(5):415–426

Dawkins R (1982) The extended phenotype. Oxford University Press, Oxford

Derks EM, Dolan CV, Boomsma DI (2006) A test of the equal environment assumption (EEA) in multivariate twin studies. Twin Res Hum Genet 9(3):403–411

Dominicus A, Skrondal A, Gjessing HK, Pedersen NL, Palmgren J (2006) Likelihood ratio tests in behavioral genetics: problems and solutions. Behav Genet 36(2):331–340

Eaves L (1976) A model for sibling effects in man. Heredity 36(2):205–214

Eaves L, Foley D, Silberg J (2003) Has the "equal environments" assumption been tested in twin studies? Twin Res Hum Genet 6(6):486–489

Eriksson M, Rasmussen F, Tynelius P (2006) Genetic factors in physical activity and the equal environment assumption: the Swedish young male twins study. Behav Genet 36(2):238–247

Evans DM, Martin NG (2000) The validity of twin studies. Gene Screen 1(2):77–79

Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, Oppermann U, Dilthey A, Pirinen M, Stone MA, Appleton L, Moutsianas L, Leslie S, Wordsworth T, Kenna TJ, Karaderi T, Thomas GP, Ward MM, Weisman MH, Farrar C, Bradbury LA, Danoy P, Inman RD, Maksymowych W, Gladman D, Rahman P, Morgan A, Marzo-Ortega H, Bowness P, Gaffney K, Gaston JS, Smith M, Bruges-Armas J, Couto AR, Sorrentino R, Paladini F, Ferreira MA, Xu H, Liu Y, Jiang L, Lopez-Larrea C, Diaz-Pena R, Lopez-Vazquez A, Zayats T, Band G, Bellenguez C, Blackburn H, Blackwell JM, Bramon E, Bumpstead SJ, Casas JP, Corvin A, Craddock N, Deloukas P, Dronov S, Duncanson A, Edkins S, Freeman C, Gillman M, Gray E, Gwilliam R, Hammond N, Hunt SE, Jankowski J, Jayakumar A, Langford C, Liddle J, Markus HS, Mathew CG, McCann OT, McCarthy MI, Palmer CN, Peltonen L, Plomin R, Potter SC, Rautanen A, Ravindrarajah R, Ricketts M, Samani N, Sawcer SJ, Strange A, Trembath RC, Viswanathan AC, Waller M, Weston P, Whittaker P, Widaa S, Wood NW, McVean G, Reveille JD, Wordsworth BP, Brown MA, Donnelly P (2011) Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet 43(8):761–767

Felson J (2014) What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. Soc Sci Res 43:184–199

Fosse R, Joseph J, Richardson K (2015) A critical assessment of the equal-environment assumption of the twin method for schizophrenia. Front Psychiatry. https://doi.org/10.3389/fpsyt.2015.00062

Genetic Analysis of Psoriasis C, the Wellcome Trust Case, Control C, Strange A, Capon F, Spencer CC, Knight J, Weale ME, Allen MH, Barton A, Band G, Bellenguez C, Bergboer JG, Blackwell JM, Bramon E, Bumpstead SJ, Casas JP, Cork MJ, Corvin A, Deloukas P, Dilthey A, Duncanson A, Edkins S, Estivill X, Fitzgerald O, Freeman C, Giardina E, Gray E, Hofer A, Huffmeier U, Hunt SE, Irvine AD, Jankowski J, Kirby B, Langford C, Lascorz J, Leman J, Leslie S, Mallbris L, Markus HS, Mathew CG, McLean WH, McManus R, Mossner R, Moutsianas L, Naluai AT, Nestle FO, Novelli G, Onoufriadis A, Palmer CN, Perricone C, Pirinen M, Plomin R, Potter SC, Pujol RM, Rautanen A, Riveira-Munoz E, Ryan AW, Salmhofer W, Samuelsson L, Sawcer SJ, Schalkwijk J, Smith CH, Stahle M, Su Z, Tazi-Ahnini R, Traupe H, Viswanathan AC, Warren RB, Weger W, Wolk K, Wood N, Worthington J, Young HS, Zeeuwen PL, Hayday A, Burden AD, Griffiths CE, Kere J, Reis A, McVean G, Evans DM, Brown MA, Barker JN, Peltonen L, Donnelly P, Trembath RC (2010) A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat Genet 42(11):985–990

Gunderson EP, Tsai AL, Selby JV, Caan B, Mayer-Davis EJ, Risch N (2006) Twins of mistaken zygosity (TOMZ): evidence for genetic contributions to dietary patterns and physiologic traits. Twin Res Hum Genet 9(4):540–549

Heath AC, Martin NG, Eaves LJ, Loesch D (1984) Evidence for polygenic epistatic interactions in man? Genetics 106(4):719–727

Heath AC, Neale MC, Hewitt JK, Eaves LJ, Fulker DW (1989) Testing structural equation models for twin data using LISREL. Behav Genet 19(1):9–35

Hemani G, Yang J, Vinkhuyzen A, Powell JE, Willemsen G, Hottenga JJ, Abdellaoui A, Mangino M, Valdes AM, Medland SE, Madden PA, Heath AC, Henders AK, Nyholt DR, de Geus EJ, Magnusson PK, Ingelsson E, Montgomery GW, Spector TD, Boomsma DI, Pedersen NL, Martin NG, Visscher PM (2013) Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. Am J Hum Genet 93(5):865–875

Hettema JM, Neale MC, Kendler KS (1995) Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders. Behav Genet 25(4):327–335

John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC (2004) Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. Am J Hum Genet 75(1):54–64

Joseph J (1998) The equal environment assumption of the classical twin method: a critical analysis. J Mind Behav 19:325–358

Joseph J (2014) The trouble with twin studies: a reassessment of twin research in the social and behavioral sciences. Routledge, London

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1993) Panic disorder in women: a population-based twin study. Psychol Med 23(2):397–406

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1993) A test of the equal-environment assumption in twin studies of psychiatric illness. Behav Genet 23(1):21–27

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (1994) Parental treatment and the equal environment assumption in twin studies of psychiatric illness. Psychol Med 24(3):579–590

Koenig LB, Jacob T, Haber JR, Xian H (2010) Testing the equal environments assumption in the children of twins design. Behav Genet 40(4):533–541

Littvay L (2012) Do heritability estimates of political phenotypes suffer from an equal environment assumption violation? Evidence from an empirical study. Twin Res Hum Genet 15(1):6–14

Loehlin JC, Nichols RC (1976) Heredity, environment, and personality: a study of 850 sets of twins. University of Texas Press, Texas

LoParo D, Waldman I (2014) Twins' rearing environment similarity and childhood externalizing disorders: a test of the equal environments assumption. Behav Genet 44(6):606–613

Lytton H (1977) Do parents create, or respond to, differences in twins? Dev Psychol 13(5):456

Maes HH, Neale MC, Eaves LJ (1997) Genetic and environmental factors in relative body weight and human adiposity. Behav Genet 27(4):325–351

Maher B (2008) Personal genomes: the case of the missing heritability. Nature 456(7218):18–21

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26(22):2867–2873

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. Nature 461(7265):747–753

Martin N, Boomsma D, Machin G (1997) A twin-pronged attack on complex traits. Nat Genet 17(4):387–392

Martin NG, Eaves LJ, Kearsey MJ, Davies P (1978) The power of the classical twin study. Heredity 40(1):97–116

Mitchell KS, Mazzeo SE, Bulik CM, Aggen SH, Kendler KS, Neale MC (2007) An investigation of a measure of twins' equal environments. Twin Res Hum Genet 10(6):840–847

Neale MC, Hunter MD, Pritikin JN, Zahery M, Brick TR, Kirkpatrick RM, Estabrook R, Bates TC, Maes HH, Boker SM (2016) OpenMx 2.0: extended structural equation and statistical modeling. Psychometrika 81(2):535–549

Neale MC, Kendler KS (1995) Models of comorbidity for multifactorial disorders. Am J Hum Genet 57(4):935–953

Nolte IM, Jansweijer JA, Riese H, Asselbergs FW, Van Der Harst P, Spector TD, Pinto YM, Snieder H, Jamshidi Y (2017) A comparison of heritability estimates by classical twin modeling and based on genome-wide genetic relatedness for cardiac conduction traits. Twin Res Hum Genet 20(6):489–498

Phillips D (1993) Twin studies in medical research: can they tell us whether diseases are genetically determined? . Lancet 341(8851):1008

Plomin R, Willerman L, Loehlin JC (1976) Resemblance in appearance and the equal environments assumption in twin studies of personality traits. Behav Genet 6(1):43–52

Posner SF, Baker L, Heath A, Martin NG (1996) Social contact, social attitudes, and twin similarity. Behav Genet 26(2):123–133

Purcell S (2002) Variance components models for gene-environment interaction in twin analysis. Twin Res 5(6):554–571

Scarr S (1968) Environmental bias in twin studies. Eugen Q 15(1):34–40

Truett KR, Eaves LJ, Walters EE, Heath AC, Hewitt JK, Meyer JM, Silberg J, Neale MC, Martin NG, Kendler KS (1994) A model system for analysis of family resemblance in extended kinships of twins. Behav Genet 24(1):35–49

Van Beijsterveldt C, Overbeek L, Rozendaal L, McMaster M, Glasner T, Bartels M, Vink J, Martin N, Dolan C, Boomsma D (2016) Chorionicity and heritability estimates from twin studies: the prenatal environment of twins and their resemblance across a large number of traits. Behav Genet 46(3):304–314

van den Oord EJ, Koot HM, Boomsma DI, Verhulst FC, Orlebeke J (1995) A twin-singleton comparison of problem behaviour in 2-3-year-olds. J Child Psychol Psychiatry 36(3):449–458

Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2(3):e41

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88(1):76–82

Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL (2013) Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet 9(5):e1003520