Incorporating polygenic scores in the twin model to estimate genotype-environment covariance: exploration of statistical power

Conor V. Dolan[1,3], Roel C. A. Huijskens, Camelia C. Minică [1], & Michael C. Neale[1,2], & Dorret I Boomsma[1]

[1]Department of Biological Psychology, Vrije Universiteit Amsterdam, Transitorium 2B03, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

[2]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, 1-156, P.O. Box 980126, Richmond, VA 23298-0126 USA

[3]Corresponding author. Email: c.v.dolan@vu.nl

Abstract

We demonstrate numerically and analytically that A-E and A-C covariance terms are identified in the univariate ACE twin model extended with polygenic scores. We conducted a power study, which showed that the power to detect these terms, given 1000 MZ and DZ twin pairs (α=0.05), depended greatly on the parameter settings of the model. We show that power is greatly improved by incorporating in the model the percentage of variance due to the polygnic scores as a fixed parameter.

Keywords

A-E covariance, A-C covariance, twin model, identification, power

The classical twin design (Eaves, et al 1978; 1977) arguably has been the most productive genetically informative design in the study of human traits (Polderman et al 2015). Multivariate and longitudinal twin studies have contributed greatly to our present knowledge concerning genetic and environmental contributions to individual differences in psychological traits (e.g., Plomin, Defries, Knopik, and Neiderhise, 2016; Turkheimer, 2002; Plomin and Deary, 2016). However, given its many strong model assumptions, results obtained with the CTD are generally viewed as provisional. The main assumptions of the CTD concern genotype-environment covariance (assumed to be absent), genotype-environment interaction (assumed to be absent), the environment assumption (assumed to be equal in MZ and DZ twin), and parental mating (assumed to be random). Given these assumptions, the CTD provides unbiased estimates of additive genetic (A), unshared environmental (E), and shared environmental (C) or dominance (D) variance components. The biasing effects of the violations of these assumption on the variance components are well understood (Verhulst & Hatemi, 2013; Purcell, 2002; Keller et al. 2009), so that estimates of variance components may be interpreted in the light of possible model violations. For instance, ignored positive A-C correlation and A-E correlation contribute to the C variance component and A variance component, respectively.

Many papers have been devoted to the detection and accommodation of model violations, either within the CTD (e.g., Purcell, 2002, Molenaar, et al 2012, Eaves & ErKanli, 2003, Carey, 1986; Dolan, et al. 2014; Beam & Turkheimer, 2013), or in extended designs (e.g., Narusyte, et al 2008; Fulker, 1988; D'Onofrio, et al 2003; Keller et al, 2009). The aim of the present paper is to demonstrate that the incorporation of polygenic scores in the classical twin design in principle allows us to estimate the covariance between A and C and between A and E. In developmental psychology, such covariances are view as plausible, stemming from processes giving rise to passive, active or evocative genotype-environment covariance (Plomin, DeFries, and McClearn, 1978; Knafo & Jaffee, 2013; Kendler, 2012; Rutter and Silberg, 2002).

The idea of incorporating measured genetic information in the twin design is not new. Previous papers concerned non-parametric linkage analysis or combined linkage-association analysis (e.g., Fulker, Cherny, Sham, Hewitt, 1999). Van den Oord and Snieder (2002) presented an extended twin model with measured genetic variable test association in the presence of stratification and to test causal relationships. In a longitudinal study of attention problems, van Beijsterveldt et al (2012) incorporated candidate genes in a common factor twin model to test the association with the latent phenotype attention problems. Minică et al (2018) presented an integration of the CTD and Mendelian randomization method, in which polygenic scores featured as genetic instruments. Bates et al. (2018) and Kong et al (2018) proposed the use of polygenic scores based on transmitted and non-transmitted alleles specifically to detect genotype-environment covariance in the genetically informative design of parents and offspring. This method may obviously be applied in the extended twin design, including twins and their parents. To the best of our knowledge, the incorporation of

polygenic scores in the twin design specifically to estimate A-E and A-C covariance has not been considered before.

The outline of this paper is as follows. First, we present the twin model, and the twin model extended with polygenic scores. Second, we address the issues of parameter identification and power using exact data simulation. Third, we present the results of our power study. We conclude this paper with a discussion.

The twin model with polygenic scores

Let Y denote the phenotypic variable of interest, and let GV_k denote the k-th genetic variant (GV) contributing to the variance of Y (k = 1 ....K). We limit ourselves to an additive model, i.e., the GVs are additively coded. The CTD can be applied to answer the question how much of the variance in Y is explained by the GV_k (k=1,…,K) in the following regression model:

$Y_i = b_0 + \sum_{k=1}^{K} b_k GV_{ki} + E^*_i$

where b_0 is the intercept, b_k is the k-th regression coefficient, subscript i denotes person, and E^* is the residual. Given a set of L of measured GVs and the complementary set M unmeasured GVs (L+M=K), we have

$Y_i = b_0 + \sum_{l=1}^{L} b_l GV_{li} + \sum_{m=1}^{M} b_m GV_{mi} + E^*_i$

Given estimates of the regression coefficients (obtained in an independent study) and the L measured GVs, the polygenic risk score equals $P_i = \sum_{l=1}^{L} b_l GV_{li}$ (Purcell et al. 2009). Assuming uncorrelated GVs, the phenotypic variance, σ^2_Y, decomposition is

$\sigma^2_Y = \sum_{l=1}^{L} b_l^2 \sigma^2_{GVl} + \sum_{m=1}^{M} b_m^2 \sigma^2_{GVm} + \sigma^2_E$.

In the classical twin design, the ACE decomposition is:

$\sigma^2_Y = \sigma^2_A + \sigma^2_C + \sigma^2_E$,

where we have partitioned the residual environmental variance σ^2_{E*} into σ^2_C and σ^2_E. Here given the presence of the polygenic scores, we can decompose the additive genetic variance σ^2_A into the observable and the latent components:

$\sigma^2_A = \sigma^2_{AL} + \sigma^2_{AK}$,

where $\sigma^2_{AL} = \sum_{l=1}^{L} b_l^2 \sigma^2_{GVl}$ (the variance of the polygenic scores, i.e., σ^2_P), and $\sigma^2_{AM} = \sum_{m=1}^{M} b_m^2 \sigma^2_{GVm}$. The information in the classical univariate twin model, i.e., the 2x2 MZ and DZ covariance matrices, is sufficient to obtain unique estimates of the variance components σ^2_A, σ^2_C, and σ^2_E. Figure 1 depicts the standard (ACE) the twin model.

The twin model including the polygenic scores is depicted in Figure 2 and 3 (to avoid clutter). Figure 2 depicts the model with A-C covariance parameter σ_{AC}, and Figure 3 depicts the model with the A-E covariance parameter σ_{AE}. The parameter R^2 is the proportion of additive genetic variance attributable to P_i ($\sum_{l=1}^{L} b_l GV_{li}$), i.e., σ^2_{AL}/σ^2_A or σ^2_P/σ^2_A. Limiting ourselves to the covariance structure, we note that the standard ACE model (Figure 1) has three parameters (variance components σ^2_A, σ^2_C, and σ^2_E), whereas the extended twin model has 6 parameters, i.e., the three variance components σ^2_A, σ^2_C, and σ^2_E, two covariance terms σ_{AE} and σ_{AC}, and the parameter R^2. The role of these parameters are shown in Table 1. Below we define effect sizes in terms of correlations, i.e., $r_{AC} = \sigma_{AC} / (\sigma_A * \sigma_C)$ and $r_{AE} = \sigma_{AE} / (\sigma_A * \sigma_E)$.

Identification and resolution

The model is identified if the observed data provide sufficient information to obtain unique estimates of the unknown parameters. In the present case, the observed information are the 3x3 MZ and the 4x4 DZ covariance matrices, containing the variances of, and the covariance among, the phenotype and the polygenic scores obtained in the twins. The MZ covariance matrix is 3x3, as MZ twins, being genetically identical, have identical polygenic scores.

As the parameterization of the phenotypic means is irrelevant to the issue of the identification of the covariance structure model, we do not consider them. Let the vector θ contain the parameters of the model:

$$\theta=[\sigma^2_A\ \sigma^2_C\ \sigma^2_E\ \sigma_{AC}\ \sigma_{AE}\ R^2].$$

The vector θ does not include σ^2_{AL} and σ^2_{AK} explicitly, because these equal $R^2*\sigma^2_A$ and $(1-R^2)*\sigma^2_A$, respectively. Local identification implies that different points in the vicinity of a given point in the admissible 6 dimensional parameter space θ are associated with the different expected covariance matrices $\Sigma_{MZ}(\theta)$ and $\Sigma_{DZ}(\theta)$ (Bollen and Bauldry, 2010; Bekker et al 1994). That is, we require that any change in the numerical value of one or more of the components in θ give rise to a change in the matrices matrices $\Sigma_{MZ}(\theta)$ and $\Sigma_{DZ}(\theta)$. We evaluated local identification analytically using symbolic algebra in Maple (Heck, 2003). The local identification check centers on the rank of the Jacobian matrix (Bekker et al 1994). The Jacobian matrix contains the first order derivatives of the elements in the matrices $\Sigma_{MZ}(\theta)$ and $\Sigma_{DZ}(\theta)$ with respect to the parameters in θ. If the Jacobian is full column rank (as established using Maple), the model is locally identified. The advantage of this method is that it is is necessary and sufficient; it does not depend on arbitrary numerical values of the parameters to establish local identification. We refer to Derks et al (2006) and Minică et al (2018) for other applications of this method in the context of twin modeling.

Having investigated local identification analytically, we proceeded to address the question of resolution by considering the statistical power to detect the parameters of interest. The issue of power is

important, as formal local identification means that we can obtain unique parameter estimates, but tells us nothing about the precision of the parameter estimates. Specifically, the model may be formally identified, but empirically underidentified. We addressed the issue of power by conducting power analyses using exact data simulation (van de Sluis et al 2008). We considered 3 parameter settings relating to the variance components, and combined these with various parameter R^2 values and covariance parameters r_{AC} and r_{AE} (see Tables 2-4). We chose R^2 = .15 and .05, which in a model with σ^2_A =.333, σ^2_C =.333, and σ^2_E =.333 (see Table 2A,B), implies that the polygenic score variance accounts for .333/.15 (~5%) or .333/.05 (~1.66%) of the phenotypic variance. We set the values of r_{AC} and r_{AE} to equal 0, .15, or .25. We considered the test of A-C covariance and A-E covariance in isolation and simultaneously in 36 additional parameter setting (see Tables 2, 3, and 4). As the effects of C are generally absent in adults (e.g., in intelligence, personality, psychopathology), we also considered the detection of σ_{AE} in an AE model (with σ^2_C=0 and r_{AC}=0). The parameter settings are shown in Table 5.

In all cases, we set the sample sizes equal to Nmz=1000 and Ndz=1000 (in total 2000 pairs), and we adopted an α of 0.05. We used normal theory maximum likelihood estimation throughout, and based our power calculations on the non-centrality parameter associated with the chi-square distribution (Martin, et al 1978). The power analyses were done using OpenMx (Boker et al 2011) in the R program (R core team, 2018). The OpenMx script is freely available (supplemental material), so that one may calculate the power in other settings.

Results

The analytical check demonstrated that the model is identified, on the basis of the 3x3 MZ and the 4x4 DZ phenotypic covariance matrices, $\Sigma_{MZ}(\theta)$ and $\Sigma_{DZ}(\theta)$. That is, we can, at least in principle, obtain unique estimates of θ=[σ^2_A σ^2_C σ^2_E σ_{AC} σ_{AE} R^2]. Table 2A, 3A, and 4A shows the power given our parameter settings (associated non-centrality parameters are given in the supplemental material). The power varies considerably in Table 2A (σ^2_A =.333, σ^2_C =.333, σ^2_E =.333), from .092 to .999. We observe power greater than the (arbitrary) value of .70 is observed in 6 cases. Of these 6, 5 are associated with the omnibus (2 df) test r_{AC}=0 <u>and</u> r_{AE} = 0, where $r_{AC} = \sigma_{AC} / (\sigma_A * \sigma_C)$ and $r_{AE} = \sigma_{AE} / (\sigma_A * \sigma_E)$. The 6th case involves the (1 df) test of r_{AE} = 0, given r_{AC} = 0 (consistent with the model in which r_{AC} = 0). Unsurprisingly, in all 6 cases, the R^2 (i.e., σ^2_P/σ^2_A) equals .15, rather than .05. In Table 3A (σ^2_A = .5, σ^2_C= .25, σ^2_E=.25), we see only 3 with power greater than .7. These again concern the 2df test (r_{AE}=0 & r_{AC}=0) and a test of r_{AE} = 0, given r_{AC} = 0 (fixed to zero, consistent with the model) in the R^2= .15 conditions. Apparently the smaller C and E variances (relative to Table 2A) result in lower power. The results in Table 4A (σ^2_A = .25, σ^2_C = .5, σ^2_E=. 25) largely resemble the results of Table 2A. As the number of

cases with power >.70 are limited in number, we conclude that the method may be underpowered given Nmz = Ndz = 1000 and α=.05.

By way of exploration, we considered the power given a known value of the parameter R^2, that is, we assume that we know how much the polygenic scores explain of the total additive genetic variance. Table 2B to 4A shows the power in the same parameter settings as given in Tables 2A to 4A, but with the parameter R^2 fixed to its true value. The power is greatly improved by this measure. In Table 2B, 3B, and 4B, we see 12, 11, and 16 cases, respectively, with power > .70. To obtain some insight into this finding, we calculated the correlations among the parameter estimates in case 2.12 (see Table 2A and 2B), i.e., given σ^2_A =.333, σ^2_C =.333, σ^2_E =.333, R^2=.15, and r_{AC} = r_{AE} = .25. We did this in the student version of LISREL 10 (Jöreskog and Sörbom, 2019), which provides the correlation matrix of the estimates based on the expected information matrix. We know from Table 2A and 2B that the power to reject r_{AC}=0 is low at .369 (Table 2A, R^2 estimated) or .561 (Table 2A, R^2 fixed), and the power to reject r_{AE}=0 is .408 (Table 2A, R^2 estimated) or .873 (Table 2B, R^2 fixed). The low power to reject r_{AC}=0 (.369) is due in part to the relatively large correlation between the parameter estimates r_{AC} and r_{AE} (-.82), implying that the effect of fixing one parameter (to zero) will be compensated by the other. We note that the the correlation between the estimates of the A and C variance terms (σ^2_A and σ^2_C) is also large (-.90), as is well-known. Having fixed the R^2 parameter, we find that the correlations among the parameters r_{AC} and r_{AE} are appreciably lower -0.61. Interestingly, the correlation between the estimates of σ^2_A and σ^2_C is only -.21, which implies that the inclusion of polygenic scores in the twin model will greatly increase the power to detect C variance. The power to detect C in the standard ACE model is known to be low in general (Visscher, Gordon, and Neale, 2008).

Table 5 shows the power to detect r_{AE} (in the absence of r_{AC}) in the AE model. With R^2 estimated, we see 7 cases (of the 20 considered) with power greater than .7. The largest power (.978) is associated with σ_A^2 = .4, σ_E^2 = .6. R^2=.15 and r_{AE} = .25. We again repeated the analyses with the R^2 parameter fixed to its true value. The power in all cases exceeds .97 (see Table 5 column 7). To provide a more insight into the variation in the power over the parameter settings considered, we repeated the analyses with Nmz=Ndz=250 (retaining α=0.05). Given Nmz=Ndz=250, we observe 10 cases with power greater than .7.

Discussion

We addressed the question of whether in incorporation of polygenic scores in the classical twin model rendered the A-C and A-E correlations (or covariances) identified. Based on the analytical results relating to the column rank of the model's matrix of partial derivatives (Jacobian) matrix and given the estimation of correct numerical values in exact data simulation (van der Sluis et al, 2008), we established that both parameters, simultaneously estimated, are identified. The power to demonstrate that these correlations are

greater than zero, in isolation (1df test) or together (2df test), given NMZ=NDZ=1000 and α=0.05, varied greatly as a function of the parameter settings. Overall, in view of the results in Table 2A to 4A, we found the likelihood ratio tests to be underpowered in many cases. Fixing the R^2 parameter to its true value helped greatly to remedy this lack of power (see Tables 2B to 4B). This solution too outlandish, as in practice, we often know the heritability of the phenotypic of interest (see Polderman et al 2015), and we know the percentage of phenotypic variance explained by the polygenic score. Based in that information, one can arrive at an approximate value of R^2. Also, in practice, one may vary the fixed R^2 values to gauge the sensitivity of the estimations of r_{AC} and / or r_{AE} to the value of R^2.

We considered a polygenic score effect size that was expressed in terms of R^2, the proportion of total additive variance attributable to the polygenic score. The values .05 and .15 are arbitrary, but not overly large. As mentioned above, these values imply that the polygenic scores account for 1.66% and 5% (respectively) of the phenotypic variance. If polygenic scores of the phenotype of interest at present do not reach the 5% or 1.66% mark, they will likely do so in the future, given the results of ever larger genome wide association meta-analyses.

We have focused on the test of r_{AC} and r_{AC} in the basic univariate twin model. Clearly the incorporation of polygenic scores may facilitate detection and modeling of r_{AC} and r_{AE} in other models. From a developmental point of view, a longitudinal study of cognitive ability would benefit great by the incorporation of polygenic scores to determine whether age-related changes in the A and C variance components of intelligence (e.g., the decrease in C variance) dependent on age-related changes in r_{AC} (Haworth et al. 2010, Tucker-Drob & Bates, 2016). While we have focused on covariance, we note that polygenetic scores have wider applications in the study of genotype-environment interplay. Polygenic scores are already being used to study interaction (e.g., Colodro-Conde et al 2018; Peyrot, et al. 2014). Assigning polygenic scores the role of moderator in the twin moderation model (Purcell, 2002) will facilitate the study of AxE and AxC interaction, allowing for the developmentally important distinction of AxC and AxE. In comparison, the method proposed by Molenaar et al (2012) is computationally more complex and likely to be less powerful, as it is based on the phenotypic twin data without polygenic scores.

In conclusion, the incorporation of polygenic scores in the twin model provides new possibilities to study interplay, as manifested in r_{AC} and r_{AE}. We found that the present approach was often underpower, given Nmz = Ndz =1000 and α=0.05, if the parameter R^2 was estimated along with the other parameters in the model. To improve power, we proposed to fix R^2 to a well-informed value (based on knowledge of the variance explained by the polygenic score). We suspect that in the coming years polygenic scores will become increasingly powerful (in terms of the amount of additive genetic variance explained) so that ultimately this "fix" will not be necessary.

References

Beam, C. R., & Turkheimer, E. (2013). Phenotype–environment correlations in longitudinal twin models. Development and Psychopathology, 25, 7–16.

Bekker, P. A., Merckens, A. & Wansbeek, T. J. (1994). Identification, Equivalent Models, and Computer Algebra: Statistical Modeling and Decision Science. Academic Press.

Boker, S.M, Neale, M.C., Hermine H. Maes, H.H, Wilde, M.J., Spiegel, M., Brick, T.R., Spies, J. Estabrook, R., Kenny, S., Bates, T.C., Mehta, P. and Fox, J. (2011). OpenMx: An Open Source Extended Structural Equation Modeling Framework. Psychometrika, 306-317. DOI:10.1007/s11336-010-9200-6

Bollen, K. A. & Bauldry, S.(2010). Model identification and computer algebra. Sociological methods & research 39 , 127-156 .

Carey G. (1986). Sibling imitation and contrast effects. Behavior Genetics, 16, 319–341. doi: 10.1007/BF01071314.

Colodro-Conde, L., Couvy-Duchesne, B., Zhu, G., Coventry, W.L., Byrne, E.M, Gordon, S., Wright, M.J., Montgomery, G.W., Madden, P.A.F., Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Ripke, S., LJ Eaves, L.J., Heath, A.C., Wray, N.R., Medland, S.E., and Martin, N.G. (2018). A direct test of the diathesis–stress model for depression Molecular Psychiatry, 23,1590–1596; doi:10.1038/mp.2017.130.

Derks, E. M., Dolan, C. V. & Boomsma, D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. Twin Research and Human Genetics 9 , 403-411.

Dolan C.V., de Kort, J.M., van Beijsterveldt, T.C., Bartels, M., and Boomsma, D.I. (2014). GE covariance through phenotype to environment transmission: an assessment in longitudinal twin data and application to childhood anxiety. Behavior Genetics.44(3), 240-253. doi: 10.1007/s10519-014-9659-5.

D'Onofrio B.M., Turkheimer E.N., Eaves L.J., Corey L.A., Berg K., Solaas M.H., et al. (2003). The role of the children of twins design in elucidating causal relations between parent characteristics and child outcomes. Journal of Child Psychology and Psychiatry, 44(8):1130–1144.

Eaves, L. J., Last, K. A., Young, P. A., and Martin, N. G. (1978). Model-fitting approaches to the analysis of human behavior.*Heredity* **41**:249–320.

Eaves LJ, Erkanli A. Markov Chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and G X E interaction. Behav Genet. 2003;33(3):279–299. doi: 10.1023/A:1023446524917.

Eaves, L. J., Last K., Martin. N. G., & Jinks J. L. (1977). A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. British Journal of Mathematical and Statistical Psychology, 30(1):1–42.

Fulker, D. W. (1988). Genetic and cultural transmission in human behavior. In Weir, B. S., Eisen, E. J., Goodman, M. M., and Namkoong, G. (eds.),*Proceedings of the Second International Conference on Quantitative Genetics*, Sinauer, Sunderland, MA, pp. 318–340.

Haworth C.M., Wright M.J., Luciano M., Martin N.G., de Geus E.J., van Beijsterveldt C.E., Bartels M., Posthuma D., Boomsma D.I., Davis O.S., Kovas Y., Corley R.P., Defries J.C., Hewitt J.K., Olson R.K., Rhea S.A., Wadsworth

S.J., Iacono W.G., McGue M., Thompson L.A., Hart S.A., Petrill S.A., Lubinski D., & Plomin R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. Molecular Psychiatry, 15(11):1112-20. doi: 10.1038/mp.2009.55.

Heath A.C., Kendler K.S., Eaves L.J., Markell D. (1985). The resolution of cultural and biological inheritance: informativeness of different relationships. Behavior Genetics, 15(5):439–465.

Heck, A. (2003). Introduction to Maple, 3rd Edition. NY: Springer-Verlag.

Keller, M.C. Medland, S.E., Duncan, L.E., Hatemi,P.K., Neale, M.C., Maes, H.H.M, and Eaves, L.J. (2009) Modeling Extended Twin Family Data I: Description of the Cascade Model. Twin Research and Human Genetics, 12(1), pp. 8–18.

Kendler, K. S. (2011). A conceptual overview of gene–environment interaction and correlation in a developmental context. In K. S. Kendler, S. R.Jaffee, & D. Romer (Eds.),The dynamic genome and mental health: The role of genes and environments in youth development(pp. 5–28). New York: Oxford University Press.

Knafo, A & Jaffee, S, R. (2013). Gene–environment correlation in developmental psychopathology Development and Psychopathology, 25, 1–6 . doi:10.1017/S0954579412000855

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., 682 Young, A. I., Thorgeirsson, T. E., … Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. Science, 359, 424–428.

Maes H.H., Neale M.C., Kendler K.S., Martin N.G., Heath A.C., & Eaves L.J. (2006). Genetic and cultural transmission of smoking initiation: an extended twin kinship model. Behavior Genetics. 2006;36(6):795–808.

Martin N.G., Eaves L.J., Kersey M.J., Davies P. The power of the classical twin study. Heredity. 1978;40:97–116. doi: 10.1038/hdy.1978.10.

Minică, C.C, Dolan, C.V., Boomsma, D.I., de Geus, E., Neale, M.C. (2018). Extending Causality Tests with Genetic Instruments: An Integration of Mendelian Randomization with the Classical Twin Design. Behavior Genetics 48:337–349.

Molenaar, D. van der Sluis, S., Boomsma, D.I., Dolan, C.V. (2012). Detecting Specific Genotype by Environment Interactions Using Marginal Maximum Likelihood Estimation in the Classical Twin Design. Behavior Genetics 42:483–499. DOI 10.1007/s10519-011-9522-x

Narusyte , J. Neiderhiser, J.M, D'Onofrio, B.M. Reiss, D. Spotts, E.L. Ganiban, J. & Lichtenstein, P. (2008).Testing different types of genotype-environment correlation: an extended children-of-twins model. Developmental Psychology 44(6), 1591–1603. doi:10.1037/a0013911.

Peyrot W.J., Milaneschi Y., Abdellaoui A., Sullivan P.F., Hottenga J.J., Boomsma D.I., et al. (2014) Effect of polygenic risk scores on depression in childhood trauma. The British Journal of Psychiatry: The Journal of Mental Science. 205(2):113–119.

Plomin R., DeFries J.C., Loehlin J.C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. Psychological Bulletin, 84:309–322. doi: 10.1037/0033-2909.84.2.309.

Plomin, R. (2014). Genotype-Environment Correlation in the Era of DNA. Behavior Genetics 44, 629–638, DOI 10.1007/s10519-014-9673-7.

Plomin,R., DeFries,J.C., Knopik, V.S., Neiderhise, J.M. (2016). Top 10 Replicated Findings from Behavioral Genetics. Perspectives in Psychological Science, 11(1): 3–23. doi:10.1177/1745691615617439

Polderman, T.J.C., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A, Visscher, P. M., Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nature Genetics, 47, 702–709.

Purcell, S. (2002) Variance components models for gene–environmentinteraction in twin analysis. Twin Res 5, 554–571.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C.,Sullivan, P. F.,...Scolnick, E. M. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature, 460(7256), 748–752. https://doi.org/10.1038/nature08185

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Tucker-Drob, E.M. and Bates, T.C. (2016). Large Cross-National Differences in Gene × Socioeconomic Status Interaction on Intelligence. Psychological Science, 27(2), 138–149.

Rutter, M., & Silberg, J. (2002). Gene–environment interplay in relation to emotional and behavioral disturbance. Annual Review of Psychology, 53, 463–490.
Turkheimer, E. (2000). Three Laws of Behavior Genetics and What They Mean. Current Directions in Psychological Sciece, 9 (5), 160 - 164.

van Beijsterveldt, C.E.M., Middeldorp, C.M., Slof-Op't Landt, M.C.T., Bartels, M, Hottenga, J.J., Eka, H., Suchiman, D, Slagboom, P.E., Boomsma, D.I. (2011). Influence of Candidate Genes on Attention Problems in Children: A Longitudinal Study. Behavior Genetics, 41:155–164. DOI 10.1007/s10519-010-9406-5

van den Oord, E.J.C.G. and Snieder, H. (2002) Including Measured Genotypes in Statistical Models to Study the Interplay of Multiple Factors Affecting Complex Traits. Behavior Genetics, 32(1), 1- 22.

van der Sluis, S., Dolan, C. V., Neale, M. C. & Posthuma, D. (2008). Power calculations using exact data simulation: a useful tool for genetic study designs. Behavior Genetics, 38 , 202-211.

Verhulst, B. and Hatemi P.K. (2013). Gene-Environment Interplay in Twin Models. Political Analysis, 21(3):368-389.

Visscher,M., Gordon, S. and Neale, M.C. (2008). Power of the Classical Twin Design Revisited: II Detection of Common Environmental Variance. Twin Research and Human Genetics, 11(1): 48–54.
doi: 10.1375/twin.11.1.48.

Table 1: Covariance terms (see Figures 2 and 3).

Covariance terms	parameterization
MZ and DZ twins	
$cov(A_{L1},E_1) = cov(A_{L2},E_2)$	$R^2\ \sigma_{AE}$
$cov(A_{M1},E_1) = cov(A_{M2},E_2)$	$(1-R^2)\ \sigma_{AE}$
$cov(A_{L1},C_1) = cov(A_{L2},C_2)$	$R^2\ \sigma_{AC}$
$cov(A_{M1},C_1) = cov(A_{M2},C_2)$	$(1-R^2)\ \sigma_{AC}$
$cov(A_{L1},C_2) = cov(A_{L2},C_1)$	$R^2\ \sigma_{AC}$
$cov(A_{M1},C_2) = cov(A_{M2},C_1)$	$(1-R^2)\ \sigma_{AC}$
MZ twins	
$cov(A_{L1},A_{L2})$	σ^2_{AL}
$cov(A_{M1},A_{M2})$	σ^2_{AM}
$cov(A_{L1},E_2) = cov(A_{L2},E_1)$	$R^2\ \sigma_{AE}$
$cov(A_{M1},E_2) = cov(A_{M2},E_1)$	$(1-R^2)\ \sigma_{AE}$
DZ twins	
$cov(A_{L1},A_{L2})$	$½*\sigma^2_{AL}$
$cov(A_{M1},A_{M2})$	$½*\sigma^2_{AM}$
$cov(A_{L1},E_2) = cov(A_{L2},E_1)$	$½\ R^2\ \sigma_{AE}$
$cov(A_{M1},E_2) = cov(A_{M2},E_1)$	$½(1-R^2)\ \sigma_{AE}$

σ^2_{AL} is the additive genetic variance due to the polygenic score
σ^2_{AM} is the residual additive genetic variance
R^2 equals $\sigma^2_{AL} / \sigma^2_A$, , where $\sigma^2_A = \sigma^2_{AL} + \sigma^2_{AM}$.
σ_{AE} and σ_{AC} are the A-E and A-C covariance terms.

Table 2A. σ^2_A =.333, σ^2_C =.333, σ^2_E =.333

set	R^2	r_{AC}	r_{AE}	r_{AC}=0 - 1 df power	r_{EC}=0 - 1df power	r_{AC}=0 & r_{AE} = 0 2 df power	r_{AC}=0 | r_{AE}=0 1 df power	r_{AE}=0 | r_{AC}=0 1df power
2.1	.05	.15	0	.092	-	.134	.092	-
2.2	.15	.15	0	.183	-	.339	.183	-
2.3	.05	.25	0	.165	-	.285	.165	-
2.4	.15	.25	0	.410	-	**.733**	.410	-
2.5	.05	0	.15	-	.093	.142	-	.183
2.6	.15	0	.15	-	.187	.366	-	.461
2.7	.05	0	.25	-	.164	.311	-	.398
2.8	.15	0	.25	**-**	.409	**.776**	**-**	**.855**
2.9	.05	.15	.15	.088	.092	.372	-	-
2.10	.15	.15	.15	.173	.186	**.858**	**-**	**-**
2.11	.05	.25	.25	.150	.164	**.761**	-	-
2.12	.15	.25	.25	.369	.408	**.999**	**-**	**-**

α = 0.05; df = degrees of freedom; Nmz=Ndz=1000. r_{AE}=0 |r_{AC}=0 denotes the test of r_{AE}=0, where r_{AC} is fixed to zero (consistent with the model). (r_{AC}=0 |r_{AE}=0 defined in the same way).

Table 2B. σ^2_A =.333, σ^2_C =.333, σ^2_E =.333

set	R^2	r_{AC}	r_{AE}	r_{AC}=0 - 1 df power	r_{EC}=0 - 1df power	r_{AC}=0 & r_{AE} = 0 2 df power	r_{AC}=0 | r_{AE}=0 1 df power	r_{AE}=0 | r_{AC}=0 1df power
2.1	.05	.15	0	.151	-	.134	.151	-
2.2	.15	.15	0	.293	-	.339	.293	-
2.3	.05	.25	0	.317	-	.285	.632	-
2.4	.15	.25	0	.632	-	**.733**	.632	-
2.5	.05	0	.15	-	.454	.429	-	.531
2.6	.15	0	.15	-	.487	.388	-	.692
2.7	.05	0	.25	-	**.848**	**.848**	**-**	**.909**
2.8	.15	0	.25	**-**	**.878**	**.956**	**-**	**.979**
2.9	.05	.15	.15	.141	.450	.624	-	-
2.10	.15	.15	.15	.268	.483	**.923**	**-**	**-**
2.11	.05	.25	.25	.280	**.842**	**.963**	-	-
2.12	.15	.25	.25	.561	**.873**	**1.00**	**-**	**-**

α = 0.05; df = degrees of freedom; Nmz=Ndz=1000, R^2 fixed to true value. r_{AE}=0 |r_{AC}=0 denotes the test of r_{AE}=0, where r_{AC} is fixed to zero (consistent with the model). (r_{AC}=0 |r_{AE}=0 defined in the same way).

Table 3A. $\sigma^2_A = .5$, $\sigma^2_C = .25$, $\sigma^2_E = .25$

set	R^2	r_{AC}	r_{AE}	$r_{AC}=0$ - 1 df power	$r_{EC}=0$ - 1df power	$r_{AC}=0$ & $r_{AE}=0$ 2 df power	$r_{AC}=0 \mid r_{AE}=0$ 1 df power	$r_{AE}=0 \mid r_{AC}=0$ 1df power
3.1	.05	.15	0	.081	-	.110	.081	-
3.2	.15	.15	0	.151	-	.262	.151	-
3.3	.05	.25	0	.136	-	.219	.136	-
3.4	.15	.25	0	.328	-	.597	.328	-
3.5	.05	0	.15	-	.082	.117	-	.148
3.6	.15	0	.15	-	.154	.284	-	.366
3.7	.05	0	.25	-	.135	.238	-	.309
3.8	.15	0	.25	-	.326	.644	-	**.743**
3.9	.05	.15	.15	.078	.082	.283	-	-
3.10	.15	.15	.15	.142	.153	**.736**	-	-
3.11	.05	.25	.25	.124	.134	.617	-	-
3.12	.15	.25	.25	.292	.325	**.988**	**-**	**-**

$\alpha = 0.05$; df = degrees of freedom; Nmz=Ndz=1000. $r_{AE}=0 \mid r_{AC}=0$ denotes the test of $r_{AE}=0$, where r_{AC} is fixed to zero (consistent with the model). ($r_{AC}=0 \mid r_{AE}=0$ defined in the same way).

Table 3B. $\sigma^2_A = .5$, $\sigma^2_C = .25$, $\sigma^2_E = .25$

set	R^2	r_{AC}	r_{AE}	$r_{AC}=0$ - 1 df power	$r_{EC}=0$ - 1df power	$r_{AC}=0$ & $r_{AE}=0$ 2 df power	$r_{AC}=0 \mid r_{AE}=0$ 1 df power	$r_{AE}=0 \mid r_{AC}=0$ 1df power
3.1	.05	.15	0	.129	-	.110	.129	-
3.2	.15	.15	0	.259	-	.262	.259	-
3.3	.05	.25	0	.260	-	.219	.260	-
3.4	.15	.25	0	.566	-	.597	.566	-
3.5	.05	0	.15	-	.526	.474	-	.578
3.6	.15	0	.15	-	.539	.583	-	.687
3.7	.05	0	.25	-	**.907**	**.888**	**-**	**.937**
3.8	.15	0	.25	-	**.916**	**.953**	**-**	**.977**
3.9	.05	.15	.15	.121	.523	.613	-	-
3.10	.15	.15	.15	.237	.536	**.874**	-	-
3.11	.05	.25	.25	.229	**.903**	**.961**	-	-
3.12	.15	.25	.25	.496	**.912**	**.999**	**-**	**-**

$\alpha = 0.05$; df = degrees of freedom; Nmz=Ndz=1000, R^2 fixed to true value. $r_{AE}=0 \mid r_{AC}=0$ denotes the test of $r_{AE}=0$, where r_{AC} is fixed to zero (consistent with the model). ($r_{AC}=0 \mid r_{AE}=0$ defined in the same way).

Table 4A. σ^2_A = .25, σ^2_C = .5, σ^2_E=. 25.

set	R^2	r_{AC}	r_{AE}	r_{AC}=0 - 1 df power	r_{EC}=0 - 1df power	r_{AC}=0 & r_{AE} = 0 2 df power	r_{AC}=0 | r_{AE}=0 1 df power	r_{AE}=0 | r_{AC}=0 1df power
4.1	.05	.15	0	.122	-	.170	.122	-
4.2	.15	.15	0	.281	-	.451	.281	-
4.3	.05	.25	0	.250	-	.384	.250	-
4.4	.15	.25	0	.620	-	**.868**	.620	-
4.5	.05	0	.15	-	.092	.123	-	.157
4.6	.15	0	.15	-	.185	.301	-	.386
4.7	.05	0	.25	-	.163	.261	-	.337
4.8	.15	0	.25	-	.405	.681	**-**	**.776**
4.9	.05	.15	.15	.117	.092	.394	-	-
4.10	.15	.15	.15	.265	.185	**.877**	**-**	**-**
4.11	.05	.25	.25	.226	.162	**.794**	**-**	**-**
4.12	.15	.25	.25	.573	.404	**.999**	**-**	**-**

α = 0.05; df = degrees of freedom; Nmz=Ndz=1000. r_{AE}=0 |r_{AC}=0 denotes the test of r_{AE}=0, where r_{AC} is fixed to zero (consistent with the model). (r_{AC}=0 |r_{AE}=0 defined in the same way).

Table 4B. σ^2_A = .25, σ^2_C = .5, σ^2_E=. 25.

set	R^2	r_{AC}	r_{AE}	r_{AC}=0 - 1 df power	r_{EC}=0 - 1df power	r_{AC}=0 & r_{AE} = 0 2 df power	r_{AC}=0 | r_{AE}=0 1 df power	r_{AE}=0 | r_{AC}=0 1df power
4.1	.05	.15	0	.199	-	.170	.199	-
4.2	.15	.15	0	.420	-	.451	.420	-
4.3	.05	.25	0	.435	-	.384	.435	-
4.4	.15	.25	0	**.815**	**-**	**.868**	**.815**	-
4.5	.05	0	.15	-	.445	.402	-	.501
4.6	.15	0	.15	-	.478	.528	-	.633
4.7	.05	0	.25	-	**.839**	**.821**	**-**	**.889**
4.8	.15	0	.25	**-**	**.871**	**.928**	**-**	**.962**
4.9	.05	.15	.15	.188	.442	.636	-	-
4.10	.15	.15	.15	.392	.475	**.933**	**-**	**-**
4.11	.05	.25	.25	**.396**	**.836**	**.968**	**-**	**-**
4.12	.15	.25	.25	**.762**	**.868**	**1.00**	**-**	**-**

α = 0.05; df = degrees of freedom; Nmz=Ndz=1000, R^2 fixed to true value. r_{AE}=0 |r_{AC}=0 denotes the test of r_{AE}=0, where r_{AC} is fixed to zero (consistent with the model). (r_{AC}=0 |r_{AE}=0 defined in the same way).

Table 5. Parameter settings, power to reject r_{AE}=0 (1 df likelihood ratio test), and the non-centrality parameter values. (α = 0.05). Column 7 and 8 "power R^2 fixed" is the power to reject r_{AE}=0, given R^2 fixed to the true value, shown in column 4.

set	σ^2_A	σ^2_E	R^2	r_{AE}	power[1]	power R^2 fixed[1]	power R^2 fixed[2]
5.1	0.3	0.7	0.05	0.15	0.336	**0.993**	0.599
5.2	0.3	0.7	0.15	0.15	**0.778**	**0.997**	0.657
5.3	0.3	0.7	0.05	0.25	0.686	**1.000**	**0.937**
5.4	0.3	0.7	0.15	0.25	**0.992**	**1.000**	**0.963**
5.5	0.4	0.6	0.05	0.15	0.287	**0.996**	0.631
5.6	0.4	0.6	0.15	0.15	**0.703**	**0.998**	0.673
5.7	0.4	0.6	0.05	0.25	0.604	**1.000**	**0.952**
5.8	0.4	0.6	0.15	0.25	**0.978**	**1.000**	**0.968**
5.9	0.5	0.5	0.05	0.15	0.243	**0.996**	0.628
5.10	0.5	0.5	0.15	0.15	0.617	**0.997**	0.663
5.11	0.5	0.5	0.05	0.25	0.519	**1.000**	**0.953**
5.12	0.5	0.5	0.15	0.25	**0.950**	**1.000**	**0.966**
5.13	0.6	0.4	0.05	0.15	0.201	**0.992**	0.592
5.14	0.6	0.4	0.15	0.15	0.522	**0.995**	0.624
5.15	0.6	0.4	0.05	0.25	0.432	**1.000**	**0.937**
5.16	0.6	0.4	0.15	0.25	**0.897**	**1.000**	**0.953**
5.17	0.7	0.3	0.05	0.15	0.162	**0.980**	0.518
5.18	0.7	0.3	0.15	0.15	0.417	**0.987**	0.551
5.19	0.7	0.3	0.05	0.25	0.342	**1.000**	**0.895**
5.20	0.7	0.3	0.15	0.25	**0.801**	**1.000**	**0.917**

[1]NMZ=NDZ=1000, [2]NMZ=NDZ=250.

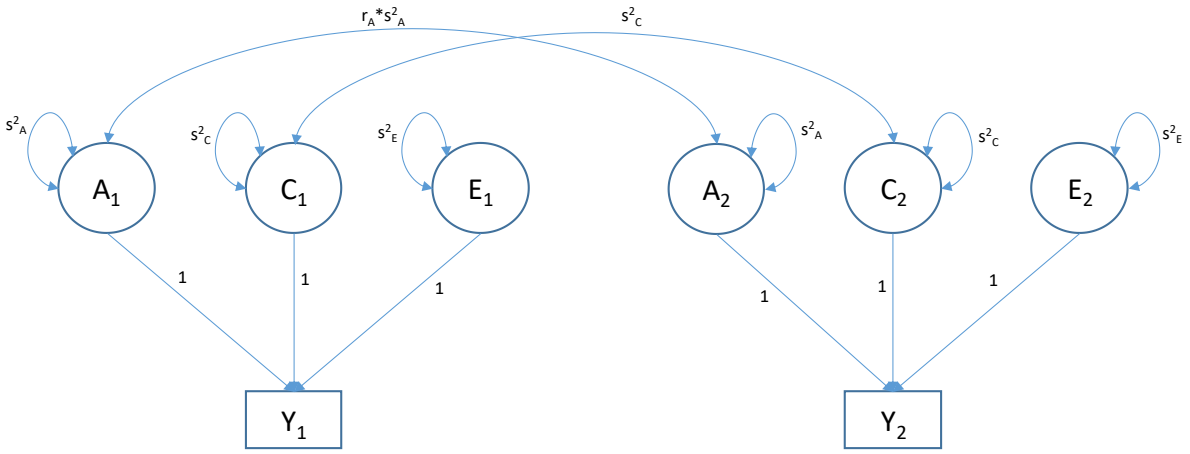

Figure 1: standard ACE twin model (r_A = 1 in MZs and r_A= .5 in DZs)

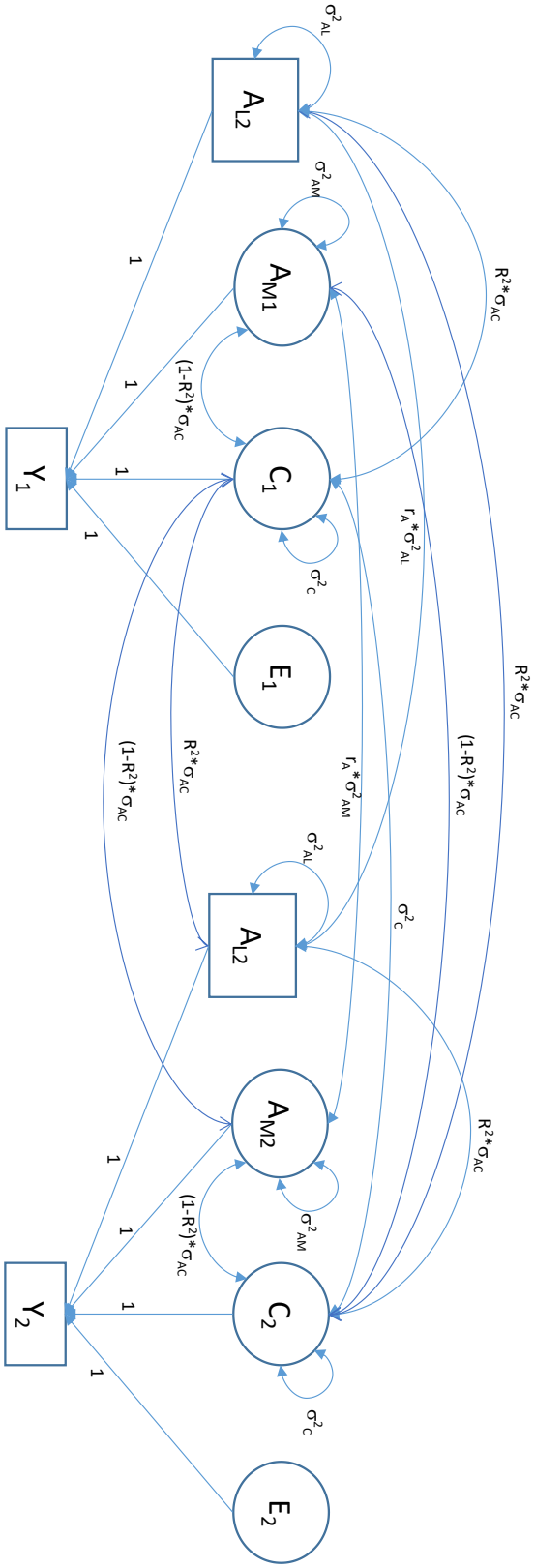

Figure 2: ACE twin model with polygenic scores (A_{L1}, A_{L2}), including A-E covariance (σ_{AE})

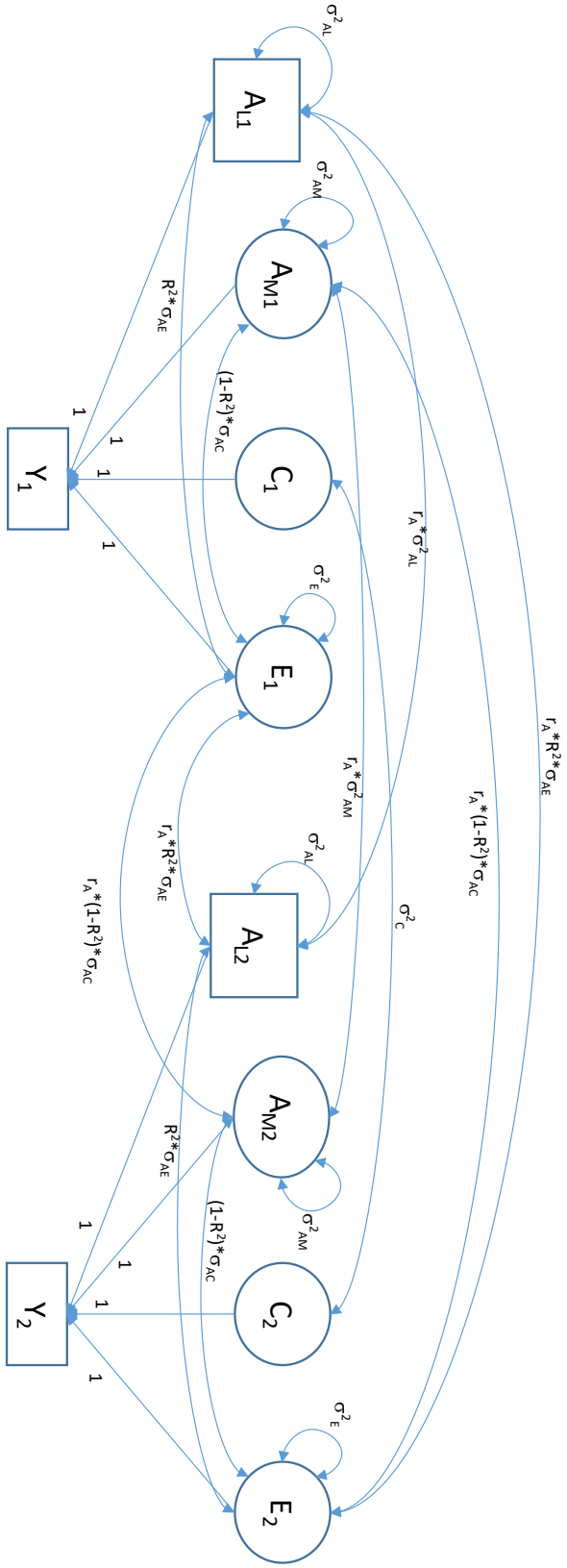

Figure 3: ACE twin model with polygenic scores (A_{L1}, A_{L2}), including A-C covariance (σ_{AC})

Supplemental material from the paper

Incorporating polygenic scores in the twin model to estimate genotype-environment covariance: exploration of statistical power

Conor V. Dolan[1,3], Roel C. A. Huijskens, Camelia C. Minică [1], & Michael C. Neale[1,2], & Dorret I Boomsma[1]

Table 2A, 3A, 4A. Non-centrality parameters (ncp2 to ncp6) Nmz=1000 Ndz=1000. pw2 to pw6 is the power (as shown in Tables).

set	vA	vC	vE	r2	cAC	cAE	df2	df3	df4	df5	df6	pw2	pw3	pw4	pw5	pw6	ncp2	ncp3	ncp4	ncp5	ncp6
2.01	0.3334	0.3333	0.3333	0.05	0.15	0.00	1	1	2	1	NA	0.092	0.050	0.134	0.092	NA	0.355	0.000	1.012	1.012	NA
2.02	0.3334	0.3333	0.3333	0.15	0.15	0.00	1	1	2	1	NA	0.183	0.050	0.339	0.183	NA	1.107	0.000	3.179	3.179	NA
2.03	0.3334	0.3333	0.3333	0.05	0.25	0.00	1	1	2	1	NA	0.165	0.050	0.285	0.165	NA	0.960	0.000	2.620	2.620	NA
2.04	0.3334	0.3333	0.3333	0.15	0.25	0.00	1	1	2	1	NA	0.410	0.050	0.733	0.410	NA	2.998	0.000	8.273	8.273	NA
2.05	0.3334	0.3333	0.3333	0.05	0.00	0.15	1	1	2	NA	1	0.050	0.093	0.142	NA	0.183	0.000	0.363	1.107	NA	1.107
2.06	0.3334	0.3333	0.3333	0.15	0.00	0.15	1	1	2	NA	1	0.050	0.187	0.366	NA	0.461	0.000	1.136	3.471	NA	3.471
2.07	0.3334	0.3333	0.3333	0.05	0.00	0.25	1	1	2	NA	1	0.050	0.164	0.311	NA	0.398	0.000	0.953	2.894	NA	2.894
2.08	0.3334	0.3333	0.3333	0.15	0.00	0.25	1	1	2	NA	1	0.050	0.409	0.776	NA	0.855	0.000	2.994	9.117	NA	9.117
2.09	0.3334	0.3333	0.3333	0.05	0.15	0.15	1	1	2	NA	NA	0.088	0.092	0.372	NA	NA	0.325	0.362	3.534	NA	NA
2.10	0.3334	0.3333	0.3333	0.15	0.15	0.15	1	1	2	NA	NA	0.173	0.186	0.858	NA	NA	1.020	1.132	11.166	NA	NA
2.11	0.3334	0.3333	0.3333	0.05	0.25	0.25	1	1	2	NA	NA	0.150	0.164	0.761	NA	NA	0.834	0.947	8.796	NA	NA
2.12	0.3334	0.3333	0.3333	0.15	0.25	0.25	1	1	2	NA	NA	0.369	0.408	0.999	NA	NA	2.640	2.980	28.027	NA	NA
3.01	0.5000	0.2500	0.2500	0.05	0.15	0.00	1	1	2	1	NA	0.081	0.050	0.110	0.081	NA	0.267	0.000	0.747	0.747	NA
3.02	0.5000	0.2500	0.2500	0.15	0.15	0.00	1	1	2	1	NA	0.151	0.050	0.262	0.151	NA	0.847	0.000	2.384	2.384	NA
3.03	0.5000	0.2500	0.2500	0.05	0.25	0.00	1	1	2	1	NA	0.136	0.050	0.219	0.136	NA	0.720	0.000	1.927	1.927	NA
3.04	0.5000	0.2500	0.2500	0.15	0.25	0.00	1	1	2	1	NA	0.328	0.050	0.597	0.328	NA	2.289	0.000	6.179	6.179	NA
3.05	0.5000	0.2500	0.2500	0.05	0.00	0.15	1	1	2	NA	1	0.050	0.082	0.117	NA	0.148	0.000	0.273	0.819	NA	0.819
3.06	0.5000	0.2500	0.2500	0.15	0.00	0.15	1	1	2	NA	1	0.050	0.154	0.284	NA	0.366	0.000	0.868	2.611	NA	2.611
3.07	0.5000	0.2500	0.2500	0.05	0.00	0.25	1	1	2	NA	1	0.050	0.135	0.238	NA	0.309	0.000	0.713	2.135	NA	2.135
3.08	0.5000	0.2500	0.2500	0.15	0.00	0.25	1	1	2	NA	1	0.050	0.326	0.644	NA	0.743	0.000	2.277	6.834	NA	6.834
3.09	0.5000	0.2500	0.2500	0.05	0.15	0.15	1	1	2	NA	NA	0.078	0.082	0.283	NA	NA	0.243	0.272	2.598	NA	NA
3.10	0.5000	0.2500	0.2500	0.15	0.15	0.15	1	1	2	NA	NA	0.142	0.153	0.736	NA	NA	0.775	0.864	8.334	NA	NA
3.11	0.5000	0.2500	0.2500	0.05	0.25	0.25	1	1	2	NA	NA	0.124	0.134	0.617	NA	NA	0.621	0.709	6.439	NA	NA
3.12	0.5000	0.2500	0.2500	0.15	0.25	0.25	1	1	2	NA	NA	0.292	0.325	0.988	NA	NA	1.994	2.265	20.791	NA	NA
4.01	0.2500	0.5000	0.2500	0.05	0.15	0.00	1	1	2	1	NA	0.122	0.050	0.170	0.122	NA	0.611	0.000	1.414	1.414	NA
4.02	0.2500	0.5000	0.2500	0.15	0.15	0.00	1	1	2	1	NA	0.281	0.050	0.451	0.281	NA	1.899	0.000	4.397	4.397	NA
4.03	0.2500	0.5000	0.2500	0.05	0.25	0.00	1	1	2	1	NA	0.250	0.050	0.384	0.250	NA	1.645	0.000	3.665	3.665	NA
4.04	0.2500	0.5000	0.2500	0.15	0.25	0.00	1	1	2	1	NA	0.620	0.050	0.868	0.620	NA	5.126	0.000	11.466	11.466	NA
4.05	0.2500	0.5000	0.2500	0.05	0.00	0.15	1	1	2	NA	1	0.050	0.092	0.123	NA	0.157	0.000	0.359	0.895	NA	0.895
4.06	0.2500	0.5000	0.2500	0.15	0.00	0.15	1	1	2	NA	1	0.050	0.185	0.301	NA	0.386	0.000	1.122	2.785	NA	2.785
4.07	0.2500	0.5000	0.2500	0.05	0.00	0.25	1	1	2	NA	1	0.050	0.163	0.261	NA	0.337	0.000	0.942	2.366	NA	2.366
4.08	0.2500	0.5000	0.2500	0.15	0.00	0.25	1	1	2	NA	1	0.050	0.405	0.681	NA	0.776	0.000	2.959	7.390	NA	7.390
4.09	0.2500	0.5000	0.2500	0.05	0.15	0.15	1	1	2	NA	NA	0.117	0.092	0.394	NA	NA	0.566	0.358	3.770	NA	NA
4.10	0.2500	0.5000	0.2500	0.15	0.15	0.15	1	1	2	NA	NA	0.265	0.185	0.877	NA	NA	1.770	1.119	11.790	NA	NA
4.11	0.2500	0.5000	0.2500	0.05	0.25	0.25	1	1	2	NA	NA	0.226	0.162	0.794	NA	NA	1.456	0.938	9.496	NA	NA
4.12	0.2500	0.5000	0.2500	0.15	0.25	0.25	1	1	2	NA	NA	0.573	0.404	0.999	NA	NA	4.592	2.949	29.967	NA	NA

Table 2B, 3B, 4B. Non-centrality parameters (ncp2 to ncp6) Nmz=1000 Ndz=1000. pw2 to pw6 is the power (as shown in Tables).

set	vA	vC	vE	R2	cAC	cAE	df2	df3	df4	df5	df6	pw2	pw3	pw4	pw5	pw6	ncp2	ncp3	ncp4	ncp5	ncp6
2.01	0.3334	0.3333	0.3333	0.05	0.15	0.00	1	1	2	1	NA	0.151	0.050	0.134	0.151	NA	0.843	0.000	1.012	1.012	NA
2.02	0.3334	0.3333	0.3333	0.15	0.15	0.00	1	1	2	1	NA	0.293	0.050	0.339	0.293	NA	1.998	0.000	3.179	3.179	NA
2.03	0.3334	0.3333	0.3333	0.05	0.25	0.00	1	1	2	1	NA	0.317	0.050	0.285	0.317	NA	2.202	0.000	2.620	2.620	NA
2.04	0.3334	0.3333	0.3333	0.15	0.25	0.00	1	1	2	1	NA	0.632	0.050	0.733	0.632	NA	5.277	0.000	8.273	8.273	NA
2.05	0.3334	0.3333	0.3333	0.05	0.00	0.15	1	1	2	NA	1	0.050	0.454	0.429	NA	0.531	0.000	3.406	4.152	NA	4.152
2.06	0.3334	0.3333	0.3333	0.15	0.00	0.15	1	1	2	NA	1	0.050	0.487	0.588	NA	0.692	0.000	3.713	6.056	NA	6.056
2.07	0.3334	0.3333	0.3333	0.05	0.00	0.25	1	1	2	NA	1	0.050	0.848	0.848	NA	0.909	0.000	8.915	10.867	NA	10.867
2.08	0.3334	0.3333	0.3333	0.15	0.00	0.25	1	1	2	NA	1	0.050	0.878	0.956	NA	0.979	0.000	9.752	15.907	NA	15.907
2.09	0.3334	0.3333	0.3333	0.05	0.15	0.15	1	1	2	NA	NA	0.141	0.450	0.624	NA	NA	0.764	3.368	6.545	NA	NA
2.10	0.3334	0.3333	0.3333	0.15	0.15	0.15	1	1	2	NA	NA	0.268	0.483	0.923	NA	NA	1.795	3.674	13.722	NA	NA
2.11	0.3334	0.3333	0.3333	0.05	0.25	0.25	1	1	2	NA	NA	0.280	0.842	0.963	NA	NA	1.892	8.782	16.647	NA	NA
2.12	0.3334	0.3333	0.3333	0.15	0.25	0.25	1	1	2	NA	NA	0.561	0.873	1.000	NA	NA	4.464	9.616	34.714	NA	NA
3.01	0.5000	0.2500	0.2500	0.05	0.15	0.00	1	1	2	1	NA	0.129	0.050	0.110	0.129	NA	0.665	0.000	0.747	0.747	NA
3.02	0.5000	0.2500	0.2500	0.15	0.15	0.00	1	1	2	1	NA	0.259	0.050	0.262	0.259	NA	1.721	0.000	2.384	2.384	NA
3.03	0.5000	0.2500	0.2500	0.05	0.25	0.00	1	1	2	1	NA	0.260	0.050	0.219	0.260	NA	1.728	0.000	1.927	1.927	NA
3.04	0.5000	0.2500	0.2500	0.15	0.25	0.00	1	1	2	1	NA	0.566	0.050	0.597	0.566	NA	4.515	0.000	6.179	6.179	NA
3.05	0.5000	0.2500	0.2500	0.05	0.00	0.15	1	1	2	NA	1	0.050	0.526	0.474	NA	0.578	0.000	4.104	4.652	NA	4.652
3.06	0.5000	0.2500	0.2500	0.15	0.00	0.15	1	1	2	NA	1	0.050	0.539	0.583	NA	0.687	0.000	4.240	5.989	NA	5.989
3.07	0.5000	0.2500	0.2500	0.05	0.00	0.25	1	1	2	NA	1	0.050	0.907	0.888	NA	0.937	0.000	10.759	12.189	NA	12.189
3.08	0.5000	0.2500	0.2500	0.15	0.00	0.25	1	1	2	NA	1	0.050	0.916	0.953	NA	0.977	0.000	11.131	15.712	NA	15.712
3.09	0.5000	0.2500	0.2500	0.05	0.15	0.15	1	1	2	NA	NA	0.121	0.523	0.613	NA	NA	0.601	4.066	6.396	NA	NA
3.10	0.5000	0.2500	0.2500	0.15	0.15	0.15	1	1	2	NA	NA	0.237	0.536	0.874	NA	NA	1.540	4.201	11.681	NA	NA
3.11	0.5000	0.2500	0.2500	0.05	0.25	0.25	1	1	2	NA	NA	0.229	0.903	0.961	NA	NA	1.480	10.626	16.368	NA	NA
3.12	0.5000	0.2500	0.2500	0.15	0.25	0.25	1	1	2	NA	NA	0.496	0.912	0.999	NA	NA	3.800	10.994	29.558	NA	NA
4.01	0.2500	0.5000	0.2500	0.05	0.15	0.00	1	1	2	1	NA	0.199	0.050	0.170	0.199	NA	1.235	0.000	1.414	1.414	NA
4.02	0.2500	0.5000	0.2500	0.15	0.15	0.00	1	1	2	1	NA	0.420	0.050	0.451	0.420	NA	3.092	0.000	4.397	4.397	NA
4.03	0.2500	0.5000	0.2500	0.05	0.25	0.00	1	1	2	1	NA	0.435	0.050	0.384	0.435	NA	3.225	0.000	3.665	3.665	NA
4.04	0.2500	0.5000	0.2500	0.15	0.25	0.00	1	1	2	1	NA	0.815	0.050	0.868	0.815	NA	8.158	0.000	11.466	11.466	NA
4.05	0.2500	0.5000	0.2500	0.05	0.00	0.15	1	1	2	NA	1	0.050	0.445	0.402	NA	0.501	0.000	3.315	3.853	NA	3.853
4.06	0.2500	0.5000	0.2500	0.15	0.00	0.15	1	1	2	NA	1	0.050	0.478	0.528	NA	0.633	0.000	3.623	5.290	NA	5.290
4.07	0.2500	0.5000	0.2500	0.05	0.00	0.25	1	1	2	NA	1	0.050	0.839	0.821	NA	0.889	0.000	8.702	10.131	NA	10.131
4.08	0.2500	0.5000	0.2500	0.15	0.00	0.25	1	1	2	NA	1	0.050	0.871	0.928	NA	0.962	0.000	9.541	13.989	NA	13.989
4.09	0.2500	0.5000	0.2500	0.05	0.15	0.15	1	1	2	NA	NA	0.188	0.442	0.636	NA	NA	1.148	3.294	6.709	NA	NA
4.10	0.2500	0.5000	0.2500	0.15	0.15	0.15	1	1	2	NA	NA	0.392	0.475	0.933	NA	NA	2.841	3.601	14.280	NA	NA
4.11	0.2500	0.5000	0.2500	0.05	0.25	0.25	1	1	2	NA	NA	0.396	0.836	0.968	NA	NA	2.879	8.625	17.193	NA	NA
4.12	0.2500	0.5000	0.2500	0.15	0.25	0.25	1	1	2	NA	NA	0.762	0.868	1.000	NA	NA	7.143	9.460	36.512	NA	NA

Table 5: non-centrality parameters (ncp3). R^2 estimated. Nmz=1000 Ndz=1000.

	set	vA	vC	vE	R2	cAC	cAE	set	df3	pw3	ncp3
5.01	0.3	0	0.7	0.05		0	0.15	4.01	1	0.336	2.356
5.02	0.3	0	0.7	0.15		0	0.15	4.02	1	0.778	7.434
5.03	0.3	0	0.7	0.05		0	0.25	4.03	1	0.686	5.978
5.04	0.3	0	0.7	0.15		0	0.25	4.04	1	0.992	18.980
5.05	0.4	0	0.6	0.05		0	0.15	4.05	1	0.287	1.951
5.06	0.4	0	0.6	0.15		0	0.15	4.06	1	0.703	6.212
5.07	0.4	0	0.6	0.05		0	0.25	4.07	1	0.604	4.945
5.08	0.4	0	0.6	0.15		0	0.25	4.08	1	0.978	15.826
5.09	0.5	0	0.5	0.05		0	0.15	4.09	1	0.243	1.588
5.10	0.5	0	0.5	0.15		0	0.15	4.10	1	0.617	5.099
5.11	0.5	0	0.5	0.05		0	0.25	4.11	1	0.519	4.033
5.12	0.5	0	0.5	0.15		0	0.25	4.12	1	0.950	13.002
5.13	0.6	0	0.4	0.05		0	0.15	4.13	1	0.201	1.254
5.14	0.6	0	0.4	0.15		0	0.15	4.14	1	0.522	4.058
5.15	0.6	0	0.4	0.05		0	0.25	4.15	1	0.432	3.197
5.16	0.6	0	0.4	0.15		0	0.25	4.16	1	0.897	10.380
5.17	0.7	0	0.3	0.05		0	0.15	4.17	1	0.162	0.938
5.18	0.7	0	0.3	0.15		0	0.15	4.18	1	0.417	3.059
5.19	0.7	0	0.3	0.05		0	0.25	4.19	1	0.342	2.407
5.20	0.7	0	0.3	0.15		0	0.25	4.20	1	0.801	7.869

Table 5: non-centrality parameters (ncp3). R^2 fixed to true value. Nmz=250 Ndz=250

	set	vA	vC	vE	R2fixed	cACfixed	cAE	set	df3	pw3	ncp3
5.01	0.3	0	0.7		0.05	0	0.15	4.01	1	0.599	4.888
5.02	0.3	0	0.7		0.15	0	0.15	4.02	1	0.657	5.586
5.03	0.3	0	0.7		0.05	0	0.25	4.03	1	0.937	12.176
5.04	0.3	0	0.7		0.15	0	0.25	4.04	1	0.963	14.053
5.05	0.4	0	0.6		0.05	0	0.15	4.05	1	0.631	5.270
5.06	0.4	0	0.6		0.15	0	0.15	4.06	1	0.673	5.795
5.07	0.4	0	0.6		0.05	0	0.25	4.07	1	0.952	13.172
5.08	0.4	0	0.6		0.15	0	0.25	4.08	1	0.968	14.580
5.09	0.5	0	0.5		0.05	0	0.15	4.09	1	0.628	5.234
5.10	0.5	0	0.5		0.15	0	0.15	4.10	1	0.663	5.663
5.11	0.5	0	0.5		0.05	0	0.25	4.11	1	0.953	13.176
5.12	0.5	0	0.5		0.15	0	0.25	4.12	1	0.966	14.318
5.13	0.6	0	0.4		0.05	0	0.15	4.13	1	0.592	4.805
5.14	0.6	0	0.4		0.15	0	0.15	4.14	1	0.624	5.182
5.15	0.6	0	0.4		0.05	0	0.25	4.15	1	0.937	12.201
5.16	0.6	0	0.4		0.15	0	0.25	4.16	1	0.953	13.194
5.17	0.7	0	0.3		0.05	0	0.15	4.17	1	0.518	4.025
5.18	0.7	0	0.3		0.15	0	0.15	4.18	1	0.551	4.357
5.19	0.7	0	0.3		0.05	0	0.25	4.19	1	0.895	10.314
5.20	0.7	0	0.3		0.15	0	0.25	4.20	1	0.917	11.189

LISREL 10 syntax. Used to estimate the correlation matrix of the parameters. The variance of the polygenic scores were increased by .001 to render the input covariance matrix positive definite. The covariance matrices included were calculated using the R code.

```
title
da ng=2 no=1000 ni=4
cm fu
 1.333333 0.075 1.000000 0.075
 0.075000 0.051 0.075000 0.050    ! .05 -> .051
 1.000000 0.075 1.333333 0.075
 0.075000 0.050 0.075000 0.051    ! .05 -> .051
la
 ph1 pg1 ph2 pg2
se
 pg1 ph1 pg2 ph2 /
mo ly=fu,fi ny=4 ne=8 te=di,fi be=ze ps=sy,fi ap=4
le
 pg1 phA1 phC1 phE1 pg2 phA2 phC2 phE2   /
ma te
.001 0 .001 0   ! to correct for .05 -> .051
ma ly
1 0 0 0   0 0 0 0  ! p1
1 1 1 1   0 0 0 0  ! ph1
0 0 0 0   1 0 0 0  ! p2
0 0 0 0   1 1 1 1  ! ph2
!
!   s2A    R2     CAC     CEC
fr par(1) par(2) par(3) par(4)
!fi par(2)
!
!
fr ps 3 3 ps 4 4 ps 7 7 ps 8 8 ps 7 3
eq ps 3 3 ps 7 7 ps 7 3
eq ps 4 4 ps 8 8
st .5 ps 3 3 ps 4 4 ps 7 7 ps 8 8 ps 7 3
!
st .333 par(1)
st .15 par(2)
st .25 par(3) par(4)
!
co ps(1,1)  = par(1)*par(2)
co ps(2,2)  = par(1) - par(1)*par(2)
co ps(5,5)  = par(1)*par(2)
co ps(6,6)  = par(1) - par(1)*par(2)
co ps(5,1) =  par(1)*par(2)           ! mz
co ps(6,2) =  par(1) -  par(1)*par(2) ! mz
!
co ps(5,3) = par(3)*par(2)
co ps(5,4) = par(4)*par(2) ! mz
co ps(6,3) = par(3) - par(3)*par(2)
co ps(6,4) =  par(4)- par(4)*par(2) !mz
co ps(3,1) =  par(3)*par(2)
co ps(3,2) =  par(3) - par(3)*par(2)
co ps(7,5)  = par(3)*par(2)
co ps(7,6)  = par(3) - par(3)*par(2)
co ps(7,1)  = par(3)*par(2)
co ps(7,2)  = par(3) - par(3)*par(2)
co ps(4,1)  = par(4)*par(2) ! mz
co ps(4,2)  = par(4)- par(4)*par(2) ! mz
co ps(8,5)  = par(4)*par(2) ! mz
co ps(8,6)  = par(4)- par(4)*par(2) ! mz
co ps(8,1)  = par(4)*par(2) ! mz
co ps(8,2)  = par(4)- par(4)*par(2) !mz
!
ou ns nd=4 ad=off
title
da no=1000
cm fu
 1.333333 0.07500 0.750000 0.04375
 0.075000 0.0500 0.043750 0.02500
 0.750000 0.04375 1.333333 0.07500
 0.043750 0.02500 0.075000 0.0500
la
```

```
 ph1 pg1 ph2 pg2
se
 pg1 ph1 pg2 ph2 /
mo ly=fu,fi ny=4 ne=8 te=di,fi be=ze ps=sy,fi
le
 pg1 phA1 phC1 phE1 pg2 phA2 phC2 phE2   /
ma te
.0 0 .0 0
ma ly
1 0 0 0   0 0 0 0  ! p1
1 1 1 1   0 0 0 0  ! ph1
0 0 0 0   1 0 0 0  ! p2
0 0 0 0   1 1 1 1  ! ph2
!
fr ps 3 3 ps 4 4 ps 7 7 ps 8 8 ps 7 3
!
eq ps 1 3 3 ps  3 3
eq ps 1 4 4 ps  4 4
eq ps 1 7 7 ps  7 7
eq ps 1 8 8 ps  8 8
eq ps 1 7 3 ps  7 3
!
co ps(1,1)  = par(1)*par(2)
co ps(2,2)  = par(1) - par(1)*par(2)  !
co ps(5,5)  = par(1)*par(2)
co ps(6,6)  = par(1) - par(1)*par(2)
co ps(5,1) =  .5*par(1)*par(2) ! mz
co ps(6,2) =  .5*par(1) - .5*par(1)*par(2) ! dz

co ps(5,3) = par(3)*par(2)
co ps(5,4) = .5*par(4)*par(2) ! mz
co ps(6,3) = par(3) - par(3)*par(2)
co ps(6,4) =  .5*par(4)- .5*par(4)*par(2) !dz


co ps(3,1) =  par(3)*par(2)
co ps(3,2) =  par(3) - par(3)*par(2)
co ps(7,5)  = par(3)*par(2)
co ps(7,6)  = par(3) - par(3)*par(2)
co ps(7,1)  = par(3)*par(2)
co ps(7,2)  = par(3) - par(3)*par(2)
co ps(4,1)  = par(4)*par(2) !
co ps(4,2)  = par(4)-par(4)*par(2) !
co ps(8,5)  = par(4)*par(2) !
co ps(8,6)  = par(4)-par(4)*par(2) !
co ps(8,1)  = .5*par(4)*par(2) ! dz
co ps(8,2)  = .5*par(4)-.5*par(4)*par(2) !dz
!
ou NS ad=off rs pc
```

R code used to calculate power given the parameter settings. The file pars1.txt contain the parameter settings (see below). Note the expected covariance matrices are called S2mz and S2d. These may be used in the LISREL codel given above. To run this model with fixed R2, one can alter the following lines:

```
mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE, values=.1, label="r2", name="R2"),
```

to

```
mxMatrix(type="Full", nrow=1, ncol=1, free=FALSE, values=R2, label="r2", name="R2"),
```

where R2 is te true value

```
# ------------------------------------ start R code
rm(list=ls(all=TRUE))
library(OpenMx)
library(MASS)
#
alpha=0.05
NMZ=NDZ=1000 # number of mz and dz twin pairs
ne=8
Slatmz=matrix(0,ne,ne) # latent covariance matrix Ap,Ar,C,E
Slatdz=matrix(0,ne,ne)
#
nset=36
pars=matrix(0,nset,6)
results=matrix(0,nset,16)
# read parameter settings from pars1.txt file
pars=read.table(file='pars1.txt',header=F)
# vA vC vE R2 rAC rAE
for (iset in 1:nset) {   #  ----36 setting from pars1.txt -----------------
setnr=pars[iset,1]
results[iset,1]=setnr
vA=pars[iset,2]
vC=pars[iset,3]
vE=pars[iset,4]
R2=pars[iset,5]
rAC=pars[iset,6] # correlations total A and C
rAE=pars[iset,7]  # added 7-may
#
vA1=R2*vA   # vA1      # variance due to Ap (polygen scores)
vA2=(1-R2)*vA  #  # variance due to Ar (background)
#
# rAC=cAC/sqrt(vA*vC)  # the correlation
cAC=rAC*sqrt(vA*vC)  # covariance
cAE=rAE*sqrt(vA*vE) # added 7-may
#
cA1C=cAC*R2
cA2C=cAC*(1-R2)
cA1E=cAE*R2    # added 7-may
cA2E=cAE*(1-R2)        # added 7-may
# exact data sim. correction for denom
cNMZ=(NMZ-1)/NMZ
cNDZ=(NDZ-1)/NDZ
# create covariance matrices
#               1  2   3   4   5  6  7   8
diag(Slatmz)=c(vE,vA1,vA2,vC,vE,vA1,vA2,vC)
diag(Slatdz)=c(vE,vA1,vA2,vC,vE,vA1,vA2,vC)
#
rz=1  # mz twins
Slatmz[2,4]=Slatmz[6,8]=R2*cAC        # A1 - C
Slatmz[4,2]=Slatmz[8,6]=R2*cAC        # C-A1
Slatmz[3,4]=Slatmz[7,8]=(1-R2)*cAC # A2 - C
Slatmz[4,3]=Slatmz[8,7]=(1-R2)*cAC # C - A2
#
Slatmz[2,6]=R2*vA*rz          # A1 - A1
Slatmz[3,7]=(1-R2)*vA*rz      # A2 - A2
Slatmz[6,2]=R2*vA*rz          #
Slatmz[7,3]=(1-R2)*vA*rz
Slatmz[4,8]=vC                # C -C
Slatmz[8,4]=vC
#
#
Slatmz[6,4]=cAC*R2;
Slatmz[4,6]=cAC*R2;
Slatmz[2,8]=cAC*R2;
```

```
Slatmz[8,2]=cAC*R2;
#
Slatmz[7,4]=cAC*(1-R2);
Slatmz[4,7]=cAC*(1-R2);
Slatmz[3,8]=cAC*(1-R2);
Slatmz[8,3]=cAC*(1-R2);
# E A cov within twin
Slatmz[2,1]=Slatmz[1,2]=cAE*(R2);  # added 7 may
Slatmz[5,6]=Slatmz[6,5]=cAE*(R2);  # added 7 may
Slatmz[3,1]=Slatmz[1,3]=cAE*(1-R2);  # added 7 may
Slatmz[5,7]=Slatmz[7,5]=cAE*(1-R2);  # added 7 may
# between
Slatmz[6,1]=Slatmz[1,6]=rz*cAE*(R2);    # added 6-7-19
Slatmz[7,1]=Slatmz[1,7]=rz*cAE*(1-R2);  # added 6-7-19
Slatmz[2,5]=Slatmz[5,2]=rz*cAE*(R2);    # added 6-7-19
Slatmz[3,5]=Slatmz[5,3]=rz*cAE*(1-R2);  # added 6-7-19
#
rz=.5  # dz twins
Slatdz[2,4]=Slatdz[6,8]=R2*cAC
Slatdz[4,2]=Slatdz[8,6]=R2*cAC
Slatdz[3,4]=Slatdz[7,8]=(1-R2)*cAC
Slatdz[4,3]=Slatdz[8,7]=(1-R2)*cAC
#
Slatdz[2,6]=R2*vA*rz
Slatdz[3,7]=(1-R2)*vA*rz
Slatdz[6,2]=R2*vA*rz
Slatdz[7,3]=(1-R2)*vA*rz
Slatdz[4,8]=vC
Slatdz[8,4]=vC
#
 #
 Slatdz[6,4]=cAC*R2;
 Slatdz[4,6]=cAC*R2;
 Slatdz[2,8]=cAC*R2;
 Slatdz[8,2]=cAC*R2;
#
 Slatdz[7,4]=cAC*(1-R2);
 Slatdz[4,7]=cAC*(1-R2);
 Slatdz[3,8]=cAC*(1-R2);
 Slatdz[8,3]=cAC*(1-R2);
#
# E A cov
Slatdz[2,1]=Slatdz[1,2]=cAE*(R2);  # added 7 may
Slatdz[5,6]=Slatdz[6,5]=cAE*(R2);  # added 7 may
Slatdz[3,1]=Slatdz[1,3]=cAE*(1-R2);  # added 7 may
Slatdz[5,7]=Slatdz[7,5]=cAE*(1-R2);  # added 7 may
#
# between  EAc
Slatdz[6,1]=Slatdz[1,6]=rz*cAE*(R2);    # added 6-7-19
Slatdz[7,1]=Slatdz[1,7]=rz*cAE*(1-R2);  # added 6-7-19 may
Slatdz[5,2]=Slatdz[2,5]=rz*cAE*(R2);    # added 6-7-19
Slatdz[5,3]=Slatdz[3,5]=rz*cAE*(1-R2);  # added 6-7-19 may
#
# create phenotypic covariance matrices.
#
L=matrix(c(
1,1,1,1,0,0,0,0,
0,1,0,0,0,0,0,0,
0,0,0,0,1,1,1,1,
0,0,0,0,0,1,0,0),4,8,byrow=T)
S2mz=L%*%Slatmz%*%t(L)
S2dz=L%*%Slatdz%*%t(L)
# simulate data MZ twins and DZ twins
dmz=matrix(0,NMZ,4)
dmz[,1:3]=mvrnorm(NMZ,rep(0,3),Sigma=S2mz[1:3,1:3]/cNMZ,emp=T)
ddz=mvrnorm(NDZ,rep(0,4),Sigma=S2dz/cNDZ,emp=T)
zyg=c(rep(1,NMZ),rep(2,NDZ))
#
dat=cbind(zyg,rbind(dmz,ddz))
#                                p1 and p2 are polygenic scores
vnames=colnames(dat)=c('zyg','T1','p1','T2','p2')
dat=as.data.frame(dat)
# openmx
dmz=dat[zyg==1,2:4]
ddz=dat[zyg==2,2:5]
Filt=matrix(c(
1,0,0,0,
```

```
0,1,0,0,
0,0,1,0),3,4,byrow=T)
#
# checks
checkS=eigen(Slatmz[1:4,1:4])$values
checkrAC=rAC
checkrAE=rAE   # added 8 may
print(checkS)  # all positive please
print(checkrAC) # rAC correlation
print(checkrAE) # rAC correlation # added 8 may
#
#
part1 <- mxModel(model="PARS",
#
# Matrices
#
# 5 parameters for the covariance structures
#
 mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE, values=1,  label="vA", name="VA"),
 mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE, values=1,  label="vC", name="VC"),
 mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE, values=1,  label="vE", name="VE"),
 mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE, values=.1, label="r2", name="R2"),
 mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE, values=.1,  label="cAC", name="CAC"),
 mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE, values=.1,  label="cAE", name="CAE"),
#
# get variances
#
 mxAlgebra(VA %*% t(VA), name="SA"),
 mxAlgebra(VC %*% t(VC), name="SC"),
 mxAlgebra(VE %*% t(VE), name="SE"),
 mxAlgebra(R2*(VA %*% t(VA)), name="SA1"),
 mxAlgebra((1-R2)*(VA %*% t(VA)), name="SA2"),
 mxAlgebra(R2*(CAC), name="CA1"),
 mxAlgebra((1-R2)*(CAC), name="CA2"),
 mxAlgebra(R2*(CAE), name="CE1"),    # added 7 may
 mxAlgebra((1-R2)*(CAE), name="CE2"), # added 7 may
#
# the MZ latent covariance matrix
# Slatdz[6,4]=cAC*R2;CA1
# Slatdz[4,6]=cAC*R2;
# Slatdz[2,8]=cAC*R2;
# Slatdz[8,2]=cAC*R2;
#
# Slatdz[7,4]=cAC*(1-R2); CA2
# Slatdz[4,7]=cAC*(1-R2);
# Slatdz[3,8]=cAC*(1-R2);
# Slatdz[8,3]=cAC*(1-R2);
#
# E A cov
#Slatmz[2,1]=Slatmz[1,2]=cAE*(R2);  # added 7 may
#Slatmz[5,6]=Slatmz[6,5]=cAE*(R2);  # added 7 may
#Slatmz[3,1]=Slatmz[1,3]=cAE*(1-R2);  # added 7 may
#Slatmz[5,7]=Slatmz[7,5]=cAE*(1-R2);  # added 7 may
#
#Slatmz[6,1]=Slatmz[1,6]=rz*cAE*(R2);    # added 6-7-19
#Slatmz[7,1]=Slatmz[1,7]=rz*cAE*(1-R2);  # added 6-7-19
#Slatmz[2,5]=Slatmz[5,2]=rz*cAE*(R2);    # added 6-7-19
#Slatmz[3,5]=Slatmz[5,3]=rz*cAE*(1-R2);  # added 6-7-19
#
# E A1 A2 C  E A1 A2 C
 mxAlgebra(expression=rbind(
#       E   A1     A2    C     E  A1 A2  C           added may 7 CE1 and CE2 terms
 cbind(SE, CE1,   CE2,   0,     0, CE1, CE2, 0),          # E
 cbind(CE1, SA1, 0,   CA1,      CE1, SA1, 0, CA1),    # A1
 cbind(CE2, 0,   SA2, CA2,     CE2, 0, SA2, CA2),            # A2
 cbind( 0, CA1, CA2, SC,        0, CA1, CA2, SC),  #  C
#
 cbind( 0, CE1,   CE2,   0,    SE, CE1,    CE2,   0),       # E
 cbind( CE1, SA1, 0,   CA1,  CE1, SA1,  0,  CA1),           # A1
 cbind( CE2, 0,   SA2, CA2,  CE2, 0,   SA2, CA2),           # A2
 cbind( 0, CA1, CA2,  SC,   0, CA1, CA2,  SC) ),name='latSmz'), # C
#
# DZ latent covariance
mxAlgebra(expression=rbind(
#       E   A1   A2     C    E  A1  A2  C            added may 7 CE1 and CE2 terms
 cbind(SE, CE1,  CE2,   0,    0, .5*CE1, .5*CE2, 0),
 cbind(CE1, SA1, 0,   CA1,  .5*CE1, .5*SA1, 0, CA1),
```

```
 cbind(CE2, 0,   SA2, CA2,  .5*CE2, 0, .5*SA2, CA2),
 cbind( 0, CA1, CA2, SC,   0, CA1, CA2, SC),  #
#
 cbind( 0, .5*CE1,   .5*CE2,   0,        SE, CE1, CE2, 0),
 cbind( .5*CE1, .5*SA1, 0,   CA1,   CE1, SA1, 0, CA1),
 cbind( .5*CE2, 0,   .5*SA2, CA2,   CE2, 0, SA2, CA2),
 cbind( 0, CA1,   CA2,   SC,    0, CA1, CA2, SC) ),name='latSdz'),
#
 mxMatrix(type="Full", nrow=3, ncol=4, free=F, values=Filt,name="F"),
 mxMatrix(type="Full", nrow=4, ncol=8, free=F, values=L,name="LY"))
#
 mzModel <- mxModel(name = "MZ",
#
 mxMatrix(type="Full",    nrow=1,    ncol=3,    free=TRUE,    values=    .0,    label=c("mT","mp","mT"),
name="expMeanMz"),
 # Algebra for expected variance/covariance matrix in MZF
 mxAlgebra(expression=PARS.F%*%PARS.LY%*%PARS.latSmz%*%t(PARS.LY)%*%t(PARS.F),name="expCovMz"),
 mxData(observed=dmz, type="raw"),
# mxFIMLObjective(covariance="expCovMz",means="expMeanMz", dimnames=vnames[2:4])
 mxExpectationNormal(covariance="expCovMz", means = "expMeanMz", vnames[2:4]),
 mxFitFunctionML(),
#
#label="vA", name="VA"),
#label="vC", name="VC"),
#label="vE", name="VE"),
#label="r2", name="R2"),
#label="cAC", name="CAC"),
#label="cAE", name="CAE"), # added7 may
#
 mxCI(c('cAC', 'cAE'))
 )
#
 dzModel <- mxModel(name = "DZ",
 mxMatrix(type="Full",   nrow=1,   ncol=4,   free=TRUE,   values=   .0,   label=c("mT","mp","mT","mp"),
name="expMeanDz"),
 # Algebra for expected variance/covariance matrix in MZF
 mxAlgebra(expression=PARS.LY%*%PARS.latSdz%*%t(PARS.LY),name="expCovDz"),
 mxData(observed=ddz, type="raw"),
# mxFIMLObjective(covariance="expCovDz",means="expMeanDz", dimnames=vnames[2:5])
 mxExpectationNormal(covariance="expCovDz", means = "expMeanDz", vnames[2:5]),
 mxFitFunctionML()
 )
# fit the model
 ACModel <- mxModel(model="TwAC", mzModel, dzModel, part1,
 mxAlgebra(expression=MZ.objective + DZ.objective, name="testfit"), #MZ.objective is the automatic name
for the -2LL of mzModel
 mxFitFunctionAlgebra("testfit"),
 mxCI(c('cAC', 'cAE'))
)
#
#Run MX
 ACModelFit <- mxRun(ACModel, intervals=F)
# summary(ACModelFit)
  parests=round(summary(ACModelFit)$parameters[3:8,5],3)
  pnames=names(parests)=summary(ACModelFit)$parameters[3:8,1]
  print(c(iset,parests))
#
#drop the parameter cAC in two parameter model
#
ACModel2=omxSetParameters(ACModel,label='cAC', free=F, value=0)
ACModel2Fit <- mxRun(ACModel2)
mxCompare(ACModelFit, ACModel2Fit)
df2=mxCompare(ACModelFit, ACModel2Fit)[2,8]
ncp2=mxCompare(ACModelFit, ACModel2Fit)[2,7]
if (abs(ncp2)<.00001) {ncp2=.00001}
crit=qchisq(alpha,df2,lower=F)  # critical value given alpha given cAC = 0.
power2=pchisq(crit,df2,ncp2,lower=F)  # power
#
# drop cAE in two parameter model
#
ACModel3=omxSetParameters(ACModel,label='cAE', free=F, value=0)
ACModel3Fit <- mxRun(ACModel3)
mxCompare(ACModelFit, ACModel3Fit)
df3=mxCompare(ACModelFit, ACModel3Fit)[2,8]
ncp3=mxCompare(ACModelFit, ACModel3Fit)[2,7]
if (abs(ncp3)<.00001) {ncp3=.00001}
crit=qchisq(alpha,df3,lower=F)  # critical value given alpha given cAC = 0.
```

```
power3=pchisq(crit,df3,ncp3,lower=F)  # power
#
#
# drop cAE and cAC omnibus test
#
ACModel4=omxSetParameters(ACModel,label=c('cAC','cAE'), free=F, value=0)
ACModel4Fit <- mxRun(ACModel4)
mxCompare(ACModelFit, ACModel4Fit)
df4=mxCompare(ACModelFit, ACModel4Fit)[2,8]
ncp4=mxCompare(ACModelFit, ACModel4Fit)[2,7]
if (abs(ncp4)<.00001) {ncp4=.00001}
crit=qchisq(alpha,df4,lower=F)  # critical value given alpha given cAC = 0.
power4=pchisq(crit,df4,ncp4,lower=F)  # power
#
#
# fixed cAC to true values and test cAE
#
df5=-1
ncp5=-1
power5=-1
if (cAE==0) {
ACModel_cAE_=omxSetParameters(ACModel,label=c('cAE'), free=F, value=c(0))
ACModel_cAEFit <- mxRun(ACModel_cAE_)
ACModel5=omxSetParameters(ACModel_cAE_,label=c('cAC'), free=F, value=c(0))
ACModel5Fit <- mxRun(ACModel5)
df5=mxCompare(ACModel_cAEFit, ACModel5Fit)[2,8]
ncp5=mxCompare(ACModel_cAEFit, ACModel5Fit)[2,7]
if (abs(ncp5)<.00001) {ncp5=.00001}
crit=qchisq(alpha,df5,lower=F)  # critical value given alpha given cAC = 0.
power5=pchisq(crit,df5,ncp2,lower=F)  # power
}

df6=-1
ncp6=-1
power6=-1
if (cAC==0) {
ACModel_cAC_=omxSetParameters(ACModel,label=c('cAC'), free=F, value=c(0))
ACModel_cACFit <- mxRun(ACModel_cAC_)
ACModel6=omxSetParameters(ACModel_cAC_,label=c('cAE'), free=F, value=c(0))
ACModel6Fit <- mxRun(ACModel6)
df6=mxCompare(ACModel_cACFit, ACModel6Fit)[2,8]
ncp6=mxCompare(ACModel_cACFit, ACModel6Fit)[2,7]
if (abs(ncp6)<.00001) {ncp6=.00001}
crit=qchisq(alpha,df6,lower=F)  # critical value given alpha given cAC = 0.
power6=pchisq(crit,df6,ncp6,lower=F)  # power
}
#
results[iset,2:16]=c(df2,df3,df4,df5,df6,
          power2, power3, power4, power5, power6,
           ncp2, ncp3, ncp4, ncp5, ncp6)
#
powplot=F
if (powplot) {
# power calculation
alpha=0.05 # alpha level
ncp=mxCompare(ACModelFit, ACModel2Fit)[2,7] # noncentrality parameters
crit=qchisq(alpha,1,lower=F)  # critical value given alpha given cAC = 0.
power1=pchisq(crit,1,ncp,lower=F)  # power
#            vary sample size to see power curve.
N=NMZ+NDZ
ncp1=ncp/(N)
N1=100:1000
N1=N1*2        # assuming NMZ = NDZ
ncp2=ncp1*N1
power2=pchisq(crit,1,ncp2,lower=F)
plot(N1,power2,type='l')
lines(N,power1,type='p',lwd=5)
} # powplot
#
# fit the standard ace model to see the bias.
#
biascheck=FALSE
if (biascheck) {
#
mzdat=dat[zyg==1,c(2,4)]
dzdat=dat[zyg==2,c(2,4)]
varnames=colnames(mzdat)
```

```
#
# the model to calculate expected summary statistics
#
nphen1=1
nphen2=2
ACEsumstat  = mxModel("ACE",
#
  # Matrix expMean for expected mean vector for MZ and DZ twins
   mxMatrix(type="Full", nrow=1, ncol=2,
            free=c(T,T),values=c(0,0),label=c("me","me"),
            name="expMean"),
#
  # Matrices a, c, and e to store the a, c, and e path coefficients
   mxMatrix(type="Lower", nrow=nphen1, ncol=nphen1,
                  free=c(T), values=t(1),
                  label=c("a11"),name="a"),
   mxMatrix(type="Lower", nrow=nphen1, ncol=nphen1,
                  free=c(T), values=c(1),
                  label=c("c11"),name="c"),
   mxMatrix(type="Lower", nrow=nphen1, ncol=nphen1,
                  free=c(T), values=c(2),
                  label=c("e11"),name="e"),
#
# Matrixes A, C, and E to compute A, C, and E variance components
#
   mxAlgebra( expression=a %*% t(a), name="A"),    # a^2
   mxAlgebra( expression=c %*% t(c), name="C"),    # c^2
   mxAlgebra( expression=e %*% t(e), name="E"),    # e^2
#
# Matrix expCovMZ for expected covariance matrix for MZ twins
   mxAlgebra( expression=
           rbind( cbind(A+C+E, A+C),
                  cbind(A+C, A+C+E)),name="expCovMZ"),
# Matrix expCovMZ for expected covariance matrix for DZ twins
   mxAlgebra( expression=
              rbind( cbind(A+C+E, .5%x%A+C),
              cbind(.5%x%A+C, A+C+E)),name="expCovDZ")
                                                        )
# a model the data, the fit function (MZ)
MZmodel=mxModel("MZ",
   mxData( observed=mzdat, type="raw"),   # the data
   mxExpectationNormal(covariance="ACE.expCovMZ",
                       means = "ACE.expMean", varnames),
    mxFitFunctionML()
                                        )
# a model the data, the fit function (DZ)
DZmodel=mxModel("DZ",
   mxData( observed=dzdat, type="raw"),
   mxExpectationNormal(covariance="ACE.expCovDZ",
                       means = "ACE.expMean", varnames),
    mxFitFunctionML()
                                      )
twinACEModel1 <-  mxModel("twinACE", ACEsumstat, MZmodel, DZmodel,
                 mxAlgebra(MZ.objective + DZ.objective, name="minus2loglikelihood"),
                 mxFitFunctionAlgebra("minus2loglikelihood")
                                                        )
# fit the model
twinACEFit1 <- mxRun(twinACEModel1)
summary(twinACEFit1)
}
# if (iset>12 & iset < 25) { print(c(iset, crit,df2,ncp2)) }
# ----------------------------
}   #  iset ... nset
# -------------------------
all=cbind(pars,round(results,3))
colnames(all)[1]='set'
colnames(all)[2:7]=pnames
colnames(all)[8]='set'
colnames(all)[9:13]=c('df2','df3','df4','df5','df6')
colnames(all)[14:18]=c('pw2','pw3','pw4','pw5','pw6')
colnames(all)[19:23]=c('ncp2','ncp3','ncp4','ncp5','ncp6')
```

The file pars1.txt contain the parameter settings (input for the R script)

```
1.01  .3334  .3333  .3333 0.05 0.15 0.00
1.02  .3334  .3333  .3333 0.15 0.15 0.00
1.03  .3334  .3333  .3333 0.05 0.25 0.00
1.04  .3334  .3333  .3333 0.15 0.25 0.00
1.05  .3334  .3333  .3333 0.05 0.00 0.15
1.06  .3334  .3333  .3333 0.15 0.00 0.15
1.07  .3334  .3333  .3333 0.05 0.00 0.25
1.08  .3334  .3333  .3333 0.15 0.00 0.25
1.09  .3334  .3333  .3333 0.05 0.15 0.15
1.10  .3334  .3333  .3333 0.15 0.15 0.15
1.11  .3334  .3333  .3333 0.05 0.25 0.25
1.12  .3334  .3333  .3333 0.15 0.25 0.25
2.01  .5000  .2500  .2500 0.05 0.15 0.00
2.02  .5000  .2500  .2500 0.15 0.15 0.00
2.03  .5000  .2500  .2500 0.05 0.25 0.00
2.04  .5000  .2500  .2500 0.15 0.25 0.00
2.05  .5000  .2500  .2500 0.05 0.00 0.15
2.06  .5000  .2500  .2500 0.15 0.00 0.15
2.07  .5000  .2500  .2500 0.05 0.00 0.25
2.08  .5000  .2500  .2500 0.15 0.00 0.25
2.09  .5000  .2500  .2500 0.05 0.15 0.15
2.10  .5000  .2500  .2500 0.15 0.15 0.15
2.11  .5000  .2500  .2500 0.05 0.25 0.25
2.12  .5000  .2500  .2500 0.15 0.25 0.25
3.01  .2500  .5000  .2500 0.05 0.15 0.00
3.02  .2500  .5000  .2500 0.15 0.15 0.00
3.03  .2500  .5000  .2500 0.05 0.25 0.00
3.04  .2500  .5000  .2500 0.15 0.25 0.00
3.05  .2500  .5000  .2500 0.05 0.00 0.15
3.06  .2500  .5000  .2500 0.15 0.00 0.15
3.07  .2500  .5000  .2500 0.05 0.00 0.25
3.08  .2500  .5000  .2500 0.15 0.00 0.25
3.09  .2500  .5000  .2500 0.05 0.15 0.15
3.10  .2500  .5000  .2500 0.15 0.15 0.15
3.11  .2500  .5000  .2500 0.05 0.25 0.25
3.12  .2500  .5000  .2500 0.15 0.25 0.25
```