

Power of the Classical Twin Design Revisited

Peter M. Visscher

School of Biological Sciences, University of Edinburgh, United Kingdom

Statistical power of the classical twin design was revisited. The approximate sampling variances of a least-squares estimate of the heritability in a univariate analysis and estimate of the genetic correlation coefficient in a bivariate analysis were derived analytically for the ACE model. Statistical power to detect additive genetic variation under the ACE model was derived analytically for least-squares, goodness-of-fit and maximum likelihood-based test statistics. The noncentrality parameter for the likelihood ratio test statistic is shown to be a simple function of the MZ and DZ intraclass correlation coefficients and the proportion of MZ and DZ twin pairs in the sample. All theoretical results were validated using simulation. The derived expressions can be used to calculate power of the classical twin design in a simple and rapid manner.

Power calculations for twin designs are useful when designing experiments to estimate variance components and to test hypotheses regarding the nature of phenotypic similarity of twins. Power calculations can be performed using asymptotic theory (Lynch & Walsh, 1998; Martin et al., 1978) or computer simulation studies based upon, for example, likelihood theory (Neale et al., 1994; Neale & Maes, 2004; Posthuma & Boomsma, 2000). Martin et al. (1978) provided a comprehensive theoretical analysis of the power of the classical twin design using weighted least-squares to estimate variance components and a goodness-of-fit test to reject 'false' models. Nowadays, maximum likelihood is commonly used to estimate variance component from twins or twin families using versatile computer programs such as Mx (Neale et al., 2002). Surprisingly, most of the literature on the estimation of parameters from twin designs is of the 'black-box' category, in that no explicit equations are given for the sampling variances of the parameter estimates and for statistical power. In this study we derive simple equations to calculate the power of twin designs under the common ACE model, contrasting least-squares with maximum likelihood and goodness-of-fit tests. Equations are presented for the sampling variance of the estimate of the heritability, the proportion of variance due to common environmental effects, and the estimate of the genetic correlation coefficient in a bivariate analysis. The noncentrality parameter

for a maximum likelihood-ratio test for genetic variance is given as a simple function of the population parameters. All predictions are verified using computer simulation.

Assumptions and Notation

Throughout, we assume the commonly used ACE model, for which the phenotypic variance is partitioned in an additive genetic (A), common environmental (C) and residual environmental (E) component. The proportions of phenotypic variance due to these random effects are h^2 , c^2 and e^2 , respectively. Predictions are first made using least squares (LS), from the properties of mean squares, which are the underlying sufficient statistics in the classical twin design. Subsequently, derivations are derived for (residual) maximum likelihood. Parameters are scaled so that total phenotypic variance is 1.0. The total variance ($\text{var}(y)$) is then partitioned as, $\text{var}(y) = h^2 + c^2 + e^2 = 1$. For a bivariate analysis, a derivation is given for the sampling variance of the estimate of the genetic correlation coefficient, using least squares. Other parameterisations and analysis methods (e.g., Jinks & Fulker, 1970) were not investigated because they are not used in practice.

Theory

Univariate models

Least squares

Consider the between-pair (B) and within-pair (W) observed mean squares (MS) in the standard ANOVA Table for n pairs, where the pairs can be either dizygotic (DZ) or monozygotic (MZ)

	df	MS	E(MS)
between pairs	$n-1$	B	$2\sigma_b^2 + \sigma_w^2$
within pair	n	W	σ_w^2

Received 13 May, 2004; accepted 14 June 2004.

Address for correspondence: Peter M. Visscher, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, Scotland, UK. E-mail: peter.visscher@ed.ac.uk

The expected mean squares and between- and within-pair variances for the ACE model, when scaled by the phenotypic variance, are

	E(B)	E(W)	σ_b^2	σ_w^2
MZ pairs	$2(h^2+c^2) + (1-h^2-c^2)$	$(1-h^2-c^2)$	h^2+c^2	$(1-h^2-c^2)$
DZ pairs	$2(1/2h^2+c^2) + (1-1/2h^2-c^2)$	$(1-1/2h^2-c^2)$	$1/2h^2+c^2$	$(1-1/2h^2-c^2)$

The variance of the observed mean squares (MS) are

$$\text{var}(\text{MS}) = 2 E(\text{MS})^2/df$$

with df the degrees of freedom. Hence the variance of the estimate of the between-pair component is

$$\text{var}(\hat{\sigma}_b^2) = \text{var}((B-W)/2)$$

$$= \frac{1}{4} \left(\frac{2E(B)^2}{n-1} + \frac{2E(W)^2}{n} \right)$$

$$\approx \frac{1}{2n} (E(B)^2 + E(W)^2)$$

From the ANOVA the estimate of the intraclass correlations is calculated as

$$\hat{t} = [(B-W)/2] / [(B-W)/2 + W] = (B-W) / (B+W)$$

Doing this for m MZ pairs and n DZ pairs gives $\hat{t}_{\text{MZ}}$ and $\hat{t}_{\text{DZ}}$. A first-order approximation of the variance of these correlations is (see, e.g., Visscher, 1998; and Lynch & Walsh, 1998 for balanced one-way designs)

$$\text{var}(\hat{t}_{\text{MZ}}) \approx (1-t_{\text{MZ}})^2(1+t_{\text{MZ}})^2/m = (1-t_{\text{MZ}}^2)^2/m$$

$$\text{var}(\hat{t}_{\text{DZ}}) \approx (1-t_{\text{DZ}})^2(1+t_{\text{DZ}})^2/n = (1-t_{\text{DZ}}^2)^2/n$$

The estimates of the genetic and common environmental components, and their approximate variances are

$$\hat{b}^2 = 2(\hat{t}_{\text{MZ}} - \hat{t}_{\text{DZ}})$$

$$\text{var}(\hat{b}^2) = 4 [\text{var}(\hat{t}_{\text{MZ}}) + \text{var}(\hat{t}_{\text{DZ}})] = 4[(1-t_{\text{MZ}}^2)^2/m + (1-t_{\text{DZ}}^2)^2/n] \quad [1]$$

or, in terms of the causal components,

$$\text{var}(\hat{b}^2) = 4\{ [1 - (h^2+c^2)^2]/m + [1 - (1/2h^2+c^2)^2]/n \}$$

Similarly, the estimate and sampling variance of the proportion of variance due to common environmental effects is

$$\hat{c}^2 = 2\hat{t}_{\text{DZ}} - \hat{t}_{\text{MZ}} \text{ and}$$

$$\text{var}(\hat{c}^2) = 4\text{var}(\hat{t}_{\text{DZ}}) + \text{var}(\hat{t}_{\text{MZ}}) = 4(1-t_{\text{DZ}}^2)^2/n + (1-t_{\text{MZ}}^2)^2/m$$

Equation [1] implies that for a given total number ($N = n+m$) of twin pairs, the sampling variance of the estimate of the heritability is minimised when

$$n/m = (1-t_{\text{DZ}}^2) / (1-t_{\text{MZ}}^2)$$

The optimum proportion of MZ pairs (p_{MZ}) is

$$p_{\text{MZ}} = (1-t_{\text{MZ}}^2) / [(1-t_{\text{DZ}}^2) + (1-t_{\text{MZ}}^2)] \quad [2]$$

Except for the trivial case when $h^2 = 0$, this ratio is smaller than $1/2$. Hence, if the cost of phenotyping is limiting and many twin pairs are available for phenotyping, then an optimum design would have more DZ than MZ twin pairs if the data are analyzed using least squares. For example, for $t_{\text{MZ}} = 0.5$ and $t_{\text{DZ}} = 0.25$, $n/m = 1.25$, that is, approximately 56% DZ and 44% MZ pairs. If $t_{\text{DZ}} = 1/2 t_{\text{MZ}}$ (AE model), and the correlation is small, then $n/m \approx 1 + 1/2 h^4$. Unless the heritability is very large (> 0.50), this suggests that the optimum design is close to a 1:1 ratio of DZ and MZ pairs.

Power and sample size. For large samples, the quantity $\lambda = (h^2/SE(\hat{b}^2))$ is the expected mean test statistic of a normal test. Its square is approximately equal to the noncentrality parameter (NCP) of a chi-square test statistic. The NCP per total number of pairs (N), is, from Equation [1],

$$\text{NCP}_{\text{LS}}/N = (t_{\text{MZ}} - t_{\text{DZ}})^2 / [(1-t_{\text{MZ}}^2)^2/p_{\text{MZ}} + (1-t_{\text{DZ}}^2)^2/(1-p_{\text{MZ}})] \quad [3]$$

For a statistical test we assume that under the null hypothesis of $h^2 = 0$ ($\lambda = 0$)

$$T = \hat{b}^2/SE(\hat{b}^2) \sim N(0,1)$$

Under the alternative hypothesis, $T \sim N(\lambda, 1)$. This allows a simple prediction of power. If $z_{1-\alpha}$ is the one-sided (upper tail) threshold from a standard normal distribution corresponding to a type-I error rate of α , and β the type-II error rate, then, for a one-sided test

$$\text{Power} = 1 - \beta = \text{Prob}(x > z_{1-\alpha} - \lambda)$$

with x a standard $N(0,1)$ random variable. Alternatively we can express the required power for a given value of the heritability in terms of the MZ and DZ sample size

$$z_{\beta} = z_{1-\alpha} - \lambda, \text{ or, } \lambda = z_{1-\beta} + z_{1-\alpha}$$

Using the variance of the estimate of the heritability

$$\lambda^2 = h^4 / \text{var}(\hat{b}^2) = (z_{1-\alpha} + z_{1-\beta})^2$$

For a given proportion of MZ twins in the sample, the required total number of twins is, from Equation [3]

$$N = 4(z_{1-\alpha} + z_{1-\beta})^2 [(1-t_{\text{MZ}}^2)^2/p_{\text{MZ}} + (1-t_{\text{DZ}}^2)/(1-p_{\text{MZ}})]/h^4$$

For example, if $p_{\text{MZ}} = 1/3$, $\alpha = 0.05$, $(1-\beta) = 0.80$, $h^2 = 0.5$ and $c^2 = 0.20$, then $z_{1-\alpha} = 1.64$, $z_{1-\beta} = 0.84$ and $N = 172$ twin pairs, $n = 115$ DZ and $m = 57$ MZ pairs. The optimal design for these parameters (from Equation [2]) is $n = 103$ and $m = 66$, for a total sample size of 169 twin pairs.

Table 1

Total Number of Pairs Required for a Power of 0.95 to Reject the CE Hypothesis When it is False at a Type-I Error Rate of 0.05

True model		Martin et al. (1978)		Maximum likelihood	
		pMZ ¹		pMZ	
h ²	c ²	0.5	Optimised ²	0.5	Optimised
0.8	0.1	68	42 (0.9)	36	33 (0.63)
0.6	0.3	85	59 (0.7)	48	45 (0.61)
0.4	0.5	123	94 (0.7)	74	72 (0.59)
0.2	0.7	277	248 (0.7)	183	180 (0.56)
0.6	0.1	257	235 (0.7)	170	168 (0.55)
0.4	0.3	466	455 (0.7)	316	314 (0.54)
0.2	0.5	1449	1449 (0.5)	1005	1002 (0.52)
0.4	0.1	940	940 (0.5)	650	649 (0.52)
0.2	0.3	3268	3268 (0.5)	2281	2279 (0.51)
0.2	0.1	5110	5110 (0.5)	3574	3573 (0.51)

Note: ¹ Proportion of MZ twins among all pairs.² Lowest total number of pairs, with the proportion of MZ pairs in brackets, selected from Table 5 of Martin et al. (1978), where the proportion of MZ twins was varied from 0.1 to 0.9, in steps of 0.2.**Maximum likelihood**

Given the sufficient statistics (sums of squares within and between MZ and DZ pairs), there is a close relationship between least squares and ML estimation for balanced designs (e.g., Thompson, 1962). In Appendix A we show the residual maximum likelihood (REML) estimation for ACE and CE models for a mixture of two one-way designs and give the expected value of the likelihood-ratio test statistic per pair from the ACE and CE model

$$NCP_{ML} = \ln\{(1-t_{AVE}^2) / [(1-t_{MZ}^2)^{p_{MZ}} (1-t_{DZ}^2)^{(1-p_{MZ})}]\} \quad [4]$$

with $t_{AVE} = p_{MZ}t_{MZ} + (1-p_{MZ})t_{DZ}$, the weighted average of the two intraclass correlations. Equation [4] contains all of the information required for a power calculation using twin pairs under the ACE model. For $p_{MZ} = 1/2$, the NCP per pair becomes

$$NCP_{ML}(p_{MZ} = 1/2) = \ln\{(1-1/4(t_{MZ} + t_{DZ})^2) / [(1-t_{MZ}^2)^{0.5} (1-t_{DZ}^2)^{0.5}]\}$$

If in addition we assume the AE model ($t_{MZ} = 2t_{DZ} = h^2$), then

$$NCP_{ML}(p_{MZ} = 1/2, t_{MZ} = 2t_{DZ}) = \ln\{(1-1/4t_{MZ}^2) / [(1-t_{MZ}^2)^{0.5} (1-4t_{MZ}^2)^{0.5}]\}$$

The required sample size, for a given value of p_{MZ} is

$$N = (z_{1-\alpha} + z_{1-\beta})^2 / NCP_{ML} = (z_{1-\alpha} + z_{1-\beta})^2 / \ln\{(1-t_{AVE}^2) / [(1-t_{MZ}^2)^{p_{MZ}} (1-t_{DZ}^2)^{(1-p_{MZ})}]\}$$

For the above numerical example of $p_{MZ} = 1/3$, $N = 152$ pairs, with 102 DZ and 51 MZ pairs. The optimal design for these parameters is $p_{MZ} = 0.546$, for $N = 127$ (69 MZ and 58 DZ pairs). For $p_{MZ} = 1/2$, $h^2 = 0.1$ and $c^2 = 0.1$, $N = 9016$ pairs for a type-I error rate of 5% and power of 80%. If the test is two-sided, then the required sample size is 11,446,

consistent with the results reported by Posthuma and Boomsma (2000, Figure 2a).

The above equations for required total sample size are remarkably simple, and only require the availability of standard statistical tables and a calculator.

Bivariate Models (Least-Squares)

For bivariate analysis, the main interest is in partitioning the phenotypic covariance in underlying components, and in particular the estimation of the genetic correlation coefficient. For notation, we use X_{ij}^Z to denote a mean square or mean cross-product. X is B (between) or W (within), Z is MZ or DZ and i and j are 1 or 2. For example, B_{ij}^{DZ} is the between-pair mean cross-product for DZ twins. The least squares estimate of the genetic correlation can be written as

$$\hat{r}_g = \frac{(B_{12}^{MZ} - W_{12}^{MZ}) - (B_{12}^{DZ} - W_{12}^{DZ})}{\sqrt{[(B_{11}^{MZ} - W_{11}^{MZ}) - (B_{11}^{DZ} - W_{11}^{DZ})][(B_{22}^{MZ} - W_{22}^{MZ}) - (B_{22}^{DZ} - W_{22}^{DZ})]}} \quad [5]$$

Table 2

Asymptotic Behaviour of Test Statistics from Least Squares, Goodness-of-Fit and Likelihood Analysis

True model	Expected test statistic		
	Least squares ($h^2 = 0$, one-sided)	Goodness-of-fit (CE model)	Likelihood ratio (ACE vs. CE, one-sided)
E	0.5	2	< 0.5 ^a
CE	0.5	2	0.5
ACE (h^2 & c^2 small)	x	x+1	x
ACE (large h^2)	< x	x	> x

Note: ^a Testing for A when the true model is E produces a zero likelihood-ratio test statistic for ACE vs. CE with a probability > 0.5

with the numerator equal to twice the difference between the MZ and DZ between-pair covariance, and the denominator the square of the product of twice the difference between the MZ and DZ between-pair variance for traits 1 and 2. Note that the genetic correlation coefficient is not defined if the estimate of the genetic variance for either trait 1 or trait 2 is negative. This happens when the between-pair mean square is smaller than the within-pair mean square, which can occur with a small sample size and/or when the population value of the heritability is small. In Appendix B we derive a Taylor series approximation of the variance of the estimate of the genetic correlation.

Equations [B2] to [B4] were used to predict the SE of the genetic correlation coefficient, and results were compared to simulations. Mean squares and mean cross-products were sampled from Wishart distributions under an ACE model and intraclass correlations and genetic correlations were estimated using least squares. When one or both of the estimated heritabilities were negative, their values were used to calculate the empirical mean and SE of the estimates, but did not contribute to an empirical estimate of the genetic correlation. One hundred thousand replicates were run for all combinations of parameters that were considered. Prediction of the sampling variance of the estimates of the heritability was very close to the observed empirical variance across replicates, with proportional differences between the observed and predicted values of approximately 1%. Observed and predicted sampling variances of genetic correlations were close, with proportional differences < 10%.

For the special case of $p_{MZ} = 1/2$, and $r_g = 0$, the approximate variance of the estimate of the correlation coefficient is

$$N * \text{var}(\hat{r}_g) = \frac{4}{(h_1^2 h_2^2)} [2(1 + 2r_p^2) + t_1^{MZ} t_2^{MZ} + t_1^{DZ} t_2^{DZ}]$$

Making a further assumption that both common environmental variance components are zero and that the heritabilities for the two traits are equal (h^2), gives

$$N * \text{var}(\hat{r}_g) = 5 + \frac{8}{h^4}$$

When dividing the numerator and denominator in equation [5] by the phenotypic standard deviations of the two traits, the equation simplifies to

$$\hat{r}_g = \frac{2(\hat{r}_b^{MZ} - \hat{r}_b^{DZ})}{\hat{h}_1 \hat{h}_2} \quad [6]$$

with $\hat{r}_b^Z$ with cross-trait correlation for MZ or DZ pairs. The sampling variance of [6] can again be approximated, but now under the assumption that the phenotypic (co)variances are estimated without error. This approximation was found to be less accurate (results not shown).

Discussion

Martin et al. (1978) addressed the power of the classical twin study in detail. Their approach and results differ from those presented here, and we discuss the reasons for these differences. Martin et al. addressed the question ‘what is the required sample size to reject a model when it is false?’. In that approach, the null hypothesis is stated but the alternative hypothesis is not. Martin et al. perform a goodness-of-fit test on the four mean squares, comparing the difference in expected mean squares under the true and ‘wrong’ model. For example, if the true model is the ACE model, Martin et al. calculate the probability (or sample size) to reject a CE or AE model. In the framework of generalised linear models, their test statistic is called ‘deviance’. Martin et al. used two degrees of freedom for the deviance test in the previous example, because there are four mean squares and two parameters under the ‘wrong’ model. However, a redundant degree of freedom is fitted if we are willing to make the assumption that the total phenotypic variances in DZ and MZ pairs are equal. A test statistic calculated from the difference in deviance from fitting an ACE model (1 degree of freedom) and a CE model (2 degrees of freedom) is very similar to a likelihood-ratio test to test the hypothesis that the additive genetic component is zero. The method of analysis used by Martin et al. (1978) was iterative weighted least-squares (WLS), which was the state of the art at that time, and is very similar to maximum likelihood.

In this study, we address the question ‘what is the required sample size to reject the null hypothesis that $h^2 = 0$ when it is false?’ in the least-squares test, and ‘what is the required sample size to reject an CE model when compared to the true ACE model?’ in the maximum-likelihood test. Hence, the null and alternative hypothesis are defined specifically, so that the degrees of freedom of a test for heritability (or genetic correlation coefficient) is one. In Table 1 we compare the required sample size to reject a CE model when the true model is ACE, using Table 5 of Martin et al. (1978) and the results herein. The required sample sizes are consistently lower when the more restricted ML test is used, by about 30%. The sample sizes are lower for two reasons, (i) because the ML test is for one degree of freedom and the GOF test for two degrees of freedom and (ii) because the ML test is one-sided ($A = 0$ vs. $A > 0$) whereas the GOF test is implicitly two-sided. A summary of the relationships between the test statistics for the LS, ML and GOF method under the null and alternative hypothesis is given in Table 2.

When contrasting the least-squares and maximum-likelihood results there are some interesting similarities and differences. If the true population correlations for the twin pairs are low then the two test statistics are asymptotically equivalent. Otherwise they are different and the least-squares test is less powerful. The least-squares analysis only uses the

contrast between the two intraclass correlations, which leads to a theoretical optimum design with more DZ than MZ twins pairs to minimise the SE of the estimate of their difference. The ML test uses both the difference in intraclass correlations between MZ and DZ pairs and the lack of fit under the reduced (CE) model. For large values of A the lack of fit contributes substantially to the test statistic which is why the optimum design has more MZ than DZ twins pairs and why the ML test is more powerful.

Williams (1993) derived asymptotic sampling variances and covariances (and sampling correlations) for estimates of variance components in commonly used models in the classical twin design. These (co)variances were derived by taking the expected value of the second differentials of the likelihood function with respect to the parameters of interest. Williams (1993) illustrated an application of these calculated asymptotic (co)variances by deriving the optimum design, in terms of the number of MZ and DZ pairs, when the objective is to minimise the asymptotic variance of A in a simple AE design. Surprisingly, the results indicated that for a large value of the heritability ($h^2 > 2/3$) the optimum design was a mixture of DZ and MZ, rather than the intuitively most efficient design of 100% MZ pairs. We have not been able to replicate Williams' findings, either in theory (see Appendix C) or by simulations.

In conclusion, simple equations were derived that approximate the sampling variances and power for estimates of genetic parameters in the classical twin design. The results presented are easily extended to other models, including CE, AE and ADE models.

Acknowledgments

I thank the U.K. Biotechnology and Biological Research Council and the U.S.A. National Institute of Mental Health (grant MH19918) for support. This paper is dedicated to Douglas Falconer, one of the founding fathers of quantitative genetics, who died in February 2004. I thank Bill Hill, Ian White, Robin

Thompson and Mike Neale for helpful discussions and comments.

References

- Anderson, T. W. (1958). *Introduction to multivariate statistical analysis*. New York: John Wiley.
- Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, 73, 311–349.
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer.
- Martin, N. G., Eaves, L. J., Kearsley, M. J., & Davies P. (1978). The power of the classical twin study. *Heredity*, 40, 97–116.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2002). *Mx: Statistical modelling* (6th ed.). Richmond, VA: Department of Psychiatry.
- Neale, M. C., Eaves, L. J., & Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics*, 24, 239–257.
- Neale, M. C., & Maes, H. H. M. (2004). *Methodology for genetics studies of twins and families*. Dordrecht, The Netherlands: Kluwer Academic.
- Posthuma, D., & Boomsma, D. I. (2000). A note on the statistical power in extended twin designs. *Behavior Genetics*, 30, 147–158.
- Thompson, W. A., Jr. (1962). The problem of negative estimates of variance components. *Annals of Mathematical Statistics*, 33, 273–289.
- Visscher, P. M. (1998). On the sampling variance of intraclass correlations and genetic correlations. *Genetics*, 149, 1605–1614.
- Williams, C. J. (1993). On the covariance between parameter estimates in models of twin data. *Biometrics*, 49, 557–568.
- Wright, F. A. (1997). The phenotypic difference discards sib-pair QTL linkage information. *American Journal of Human Genetics*, 60, 740–742.

Appendix A: Residual Maximum Likelihood (REML) estimation from mean squares

Given the sum of squares (SS) about the mean for DZ and MZ pairs, the log-likelihood function (L) has the following form

$$-2L = (m-1)\ln[E(B^{MZ})] + m\ln[E(W^{MZ})] + (n-1)\ln[E(B^{DZ})] + n\ln[E(W^{DZ})] \\ + SSB^{MZ}/E(B^{MZ}) + SSW^{MZ}/E(W^{MZ}) + SSB^{DZ}/E(B^{DZ}) + SSW^{DZ}/E(W^{DZ}) \quad [A1]$$

(for example, Thompson, 1962). This is the function for the residual (or restricted) likelihood because the degree of freedom in estimating the MZ and DZ means have been taken into account. For twin analyses, this adjustment is trivial and the residual likelihood and REML estimates are very similar to the standard likelihood and ML estimates.

For the ACE model, the expected values of the 4 mean squares are functions of 3 parameters, so an iterative procedure has to be used. For the CE model, the REML estimates of the variance between (subscript *bs*) and within (subscript *ws*) strata are

$$\hat{\sigma}_{bs}^2 = [SSB^{MZ} + SSB^{DZ}] / [m+n-2], \text{ and } \hat{\sigma}_{ws}^2 = [SSW^{MZ} + SSW^{DZ}] / [m+n] \quad [A2]$$

The REML estimates of the underlying variance components are, if $\hat{\sigma}_{bs}^2 > \hat{\sigma}_{ws}^2$,

$$\hat{\sigma}_e^2 = \hat{\sigma}_{ws}^2 \text{ and } \hat{\sigma}_c^2 = (\hat{\sigma}_{bs}^2 - \hat{\sigma}_{ws}^2) / 2$$

If $\hat{\sigma}_{bs}^2 < \hat{\sigma}_{ws}^2$, then the REML estimates are $\hat{\sigma}_c^2 = 0$ and

$$\hat{\sigma}_e^2 = [SSB^{MZ} + SSB^{DZ} + SSW^{MZ} + SSW^{DZ}] / [2(m+n-1)] \\ = [(m+n-2)\hat{\sigma}_{bs}^2 + (m+n)\hat{\sigma}_{ws}^2] / [2(m+n-1)]$$

The expected log-likelihood-ratio test statistic can be approximated by replacing the SS in [A1] by the expected value under the true (ACE) model, and the expected estimates of the variance components under either the full model or the reduced model. Asymptotically, the expected variance components under the CE model are obtained from [A2], by replacing the SS by their expected values.

Asymptotically ($m \approx m-1$ and $n \approx n-1$), the log-likelihood function, scaled by the total number of twin pairs (N) can be written as

$$-2L/N = p_{MZ}\ln[E(B^{MZ})] + p_{MZ}\ln[E(W^{MZ})] + (1-p_{MZ})\ln[E(B^{DZ})] + (1-p_{MZ})\ln[E(W^{DZ})] \\ + p_{MZ}B^{MZ}/E(B^{MZ}) + p_{MZ}W^{MZ}/E(W^{MZ}) + \\ + (1-p_{MZ})B^{DZ}/E(B^{DZ}) + (1-p_{MZ})W^{DZ}/E(W^{DZ}) \quad [A3]$$

The NCP per pair of the likelihood-ratio-test (LRT) for the ACE versus the CE model was derived using [A2] and [A3]

$$NCP = \ln\{(1-t_{AVE}^2) / [(1-t_{MZ}^2)^{p_{MZ}} (1-t_{DZ}^2)^{(1-p_{MZ})}]\} \quad [A4]$$

with $t_{AVE} = p_{MZ}t_{MZ} + (1-p_{MZ})t_{DZ}$, the weighted average of the two intraclass correlations. Equation [A4] is simple and shows a strong resemblance to the NCP for QTL mapping using sibpairs (Wright, 1997). At the limit, for both t_{MZ} and t_{DZ} close to zero, equation [A4] reduces to $p_{MZ}(1-p_{MZ})(t_{MZ}-t_{DZ})^2$, which is equivalent to the limit of the NCP for least squares (Eq. [3] in the main text). To derive the optimum value of p_{MZ} , equation [A4] was differentiated with respect to p_{MZ} . The maximum NCP is achieved when

$$p_{MZ} = [X^{-2} + Y^{-2}]^{0.5} + X^{-1} - t_{DZ}Y^{-1} \quad [A5]$$

with $X = \ln[(1-t_{MZ}^2) / (1-t_{DZ}^2)]$ and $Y = (t_{MZ}-t_{DZ})$

The optimum proportion of MZ twins is always > 0.5 . For a wide range of plausible combinations of parameters it is between 0.5 and 0.6 (consistent with Martin et al., 1978).

Appendix B: Approximate sampling variance for the least square estimator of the genetic correlation coefficient

From the main text, a least-squares estimator of the estimate of the genetic correlation coefficient is

$$\hat{r}_g = \frac{(B_{12}^{MZ} - W_{12}^{MZ}) - (B_{12}^{DZ} - W_{12}^{DZ})}{\sqrt{[(B_{11}^{MZ} - W_{11}^{MZ}) - (B_{11}^{DZ} - W_{11}^{DZ})][(B_{22}^{MZ} - W_{22}^{MZ}) - (B_{22}^{DZ} - W_{22}^{DZ})]}} \quad [B1]$$

To approximate the variance of [B1], we repeatedly use the following results from standard multivariate theory (e.g., Anderson, 1958), for mean squares X_{ij}^Z and X_{kl}^Z .

$$\text{cov}(X_{ij}^Z, X_{kl}^Z) = [E(X_{ik}^Z)E(X_{jl}^Z) + E(X_{il}^Z)E(X_{jk}^Z)]/df$$

with df the degrees of freedom. We assume that the number of MZ and DZ pairs is large enough so that $n \approx n-1$ and $m \approx m-1$. For example,

$$\text{var}(B_{12}^{DZ}) = [E(B_{11}^{DZ})E(B_{22}^{DZ}) + E(B_{12}^{DZ})^2] / n$$

Covariances among between and within mean squares and cross-products are zero, as are covariances between mean squares and cross-products from MZ and DZ pairs. Equation [B1] was approximated by a number of first-order Taylor series (for example, Lynch & Walsh, 1998)

$$\text{var}(X/Y) \sim \text{var}(X)/E(Y)^2 + \text{var}(Y)E(X)^2/E(Y)^4 - 2\text{cov}(X,Y)E(X)/E(Y)^3 \quad [B2]$$

$$\text{var}(XY) \sim E(X)^2\text{var}(Y) + E(Y)^2\text{var}(X) + 2E(X)E(Y)\text{cov}(X,Y) \quad [B3]$$

and

$$\text{var}(X^{0.5}) \sim 1/4\text{var}(X)/E(X) \quad [B4]$$

with X the numerator and Y the denominator. Using first-order Taylor series results in the following expression for the mean and (co)variances of the X and Y terms:

$$E(X) = r_s h_1 h_2$$

$$E(Y) = h_1 h_2$$

$$N * \text{var}(X) = \frac{2}{p_{MZ}} [1 + t_1^{MZ} t_2^{MZ} + r_p^2 + (r_b^{MZ})^2] + \frac{2}{(1 - p_{MZ})} [1 + t_1^{DZ} t_2^{DZ} + r_p^2 + (r_b^{DZ})^2]$$

$$N * \text{var}(Y) = \frac{h_1^2}{h_2^2} \left[\frac{1 + (t_2^{MZ})^2}{p_{MZ}} + \frac{1 + (t_2^{DZ})^2}{(1 - p_{MZ})} \right] + \frac{h_2^2}{h_1^2} \left[\frac{1 + (t_1^{MZ})^2}{p_{MZ}} + \frac{1 + (t_1^{DZ})^2}{(1 - p_{MZ})} \right] + 2 \left[\frac{r_p^2 + (r_b^{MZ})^2}{p_{MZ}} + \frac{r_p^2 + (r_b^{DZ})^2}{(1 - p_{MZ})} \right]$$

and

$$N * \text{cov}(X, Y) = 2 \left(\frac{h_1}{h_2} \left[\frac{r_p + t_2^{MZ} r_b^{MZ}}{p_{MZ}} + \frac{r_p + t_2^{DZ} r_b^{DZ}}{(1 - p_{MZ})} \right] + \frac{h_2}{h_1} \left[\frac{r_p + t_1^{MZ} r_b^{MZ}}{p_{MZ}} + \frac{r_p + t_1^{DZ} r_b^{DZ}}{(1 - p_{MZ})} \right] \right)$$

Appendix C: Asymptotic sampling variances under an AE model

We considered a sample consisting of either all MZ or all DZ pairs. Taking the expectation of the 2nd differentials of the likelihood function (Appendix A) with respect to A, E and the combination AE, and inverting the resulting 2 x 2 matrix gives the asymptotic covariance matrix for the two variance components. When we scale the results by the number of pairs and by the phenotypic variance, the results become:

Asymptotic variances	MZ	DZ
$\text{var}(\hat{\sigma}_A^2)$	$1+h^4$	$4+h^4$
$\text{var}(\hat{\sigma}_E^2)$	$2(1-h^2)^2$	$\frac{1}{2}[(1+\frac{1}{2}h^2)^2 + 9(1-\frac{1}{2}h^2)^2]$
$\text{var}(\hat{\sigma}_A^2) h^2 \rightarrow 0$	1	4
$\text{var}(\hat{\sigma}_E^2) h^2 \rightarrow 0$	2	5
$\text{var}(\hat{\sigma}_A^2) h^2 \rightarrow 1$	2	5
$\text{var}(\hat{\sigma}_E^2) h^2 \rightarrow 1$	0	$\frac{9}{4}$

These results were verified by computer simulation. It is clear that per pair the asymptotic sampling variance for MZ is always smaller than that for DZ pairs, irrespective of the heritability, as one might expect. A least squares analogy gives the same result because the variances of the intraclass correlations for MZ and DZ pairs are proportional to $(1-h^2)^2$ and $(1-\frac{1}{2}h^2)^2$, respectively, so always lower per MZ pair.