Detecting Genotype–Environment Interaction in Monozygotic Twin Data: Comparing the Jinks and Fulker Test and a New Test Based on Marginal Maximum Likelihood Estimation

Sophie van der Sluis,[1] Conor V. Dolan,[2] Michael C. Neale,[3] Dorret I. Boomsma,[1] and Danielle Posthuma[1]

[1] *Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands*
[2] *Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands*
[3] *Departments of Psychiatry, Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, United States of America*

This article is concerned with the power to detect the presence of genotype by environment interaction (G × E) in the case that both genes and environment feature as latent (i.e., unmeasured) variables. The power of the test proposed by Jinks and Fulker (1970), which is based on regressing the absolute difference between the scores of monozygotic twins on the sums of these scores, is compared to the power of an alternative test, which is based on Marginal Maximum Likelihood (MML). Simulation studies showed that generally the power of the MML-based test was greater than the power of the Jinks and Fulker test in detecting linear and curvilinear G × E interaction, regardless of whether the distribution of the data deviated significantly from normality. However, after a normalizing transformation, the Jinks and Fulker test performed slightly better. Some possible future extensions of the MML-based test are briefly discussed.

The standard twin design includes a number of well-known assumptions (Boomsma et al., 2002; Eaves et al., 1989; Neale & Cardon, 1992). One important assumption is that genotype by environment interaction is absent. The presence of an interaction between genotype and either shared or unshared environment renders the usual summary statistics, such as the heritability coefficient, h^2 (broad-sense) or a^2 (narrow-sense), difficult to interpret. This is because genotype by environment interaction implies that the effect of the environment depends on an individual's genetic make-up, or that the genetic effect depends on the individual's environment (Purcell, 2002). So, while in the standard additive genetic twin model, the phenotypic effect is assumed to be the sum of additive genetic effects, and shared and unshared environmental effects (i.e., P = G + C + E, where P is the phenotype, G the genetic effect, C the shared environmental effect, and E the unshared environmental effect), one can also assume the presence of an interaction between genes and environment (e.g., phenotype = additive genetic effects + unshared environmental effects + an interaction between genotype and unshared environment; P = G + E + G × E, i.e., a model in which shared environmental effects are assumed absent). The focus of the present article is on this interaction between genotype and unshared environment (henceforth G × E).

Various methods have been considered to detect and/or model G × E. Multi-group designs can be used to test the presence of G × E when actual measures of either G or E are available. Examples include the effect of the *APOE e4* allele on cognitive decline in males versus females (Yaffe et al., 2000), and the *5-HTT* gene and the effect of stressful life events on the risk of depression (Caspi et al., 2003). Alternatively, G may feature as a latent (i.e., unmeasured) variable, and E as a measured categorical moderator. Here, examples include the interaction between marital status and genetic risk for depression (Heath et al., 1998), the interaction between past-year life events and genetic liability for depression and anxiety (Silberg et al., 2001), and the interaction between upbringing and genetic liability for disinhibition (Boomsma et al., 1999). Purcell (2002) introduced a model for testing G × E in which G features as a latent variable, and E as a continuous

Received 19 December, 2005; accepted 26 January, 2006.

Address for correspondence: Sophie van der Sluis, Department of Biological Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands. E-mail: S.van.der.Sluis@psy.vu.nl

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

moderator. This model has, for example, been used to study the effect of family dysfunction on the heritability of neuroticism (Kendler et al., 2003), and the effect of socioeconomic status on the heritability of IQ (Turkheimer et al., 2003). When both G and E feature as unmeasured, latent variables, the detection of G × E interaction is more difficult. Molenaar and Boomsma (1987) considered a detection method based on the higher order moments of genetic and environmental factor scores. This method requires multivariate data, and large sample sizes to ensure the stability of estimates of higher order moments. Jinks and Fulker (1970) suggested a detection method based on the regression of absolute differences of monozygotic (MZ) twin-pair scores (i.e., $|y_1 - y_2|$), on the sums of MZ twin-pair scores (i.e., $y_1 + y_2$). In MZ twins reared apart, the difference $|y_1 - y_2|$ can be taken as an estimate of all postnatal environmental effects. For MZ twins reared together, the difference $|y_1 - y_2|$ only functions as an estimate of the environmental effects unique to individuals within a family. The expected value of this difference score will be equal across families ($|y_{11} - y_{12}| = \ldots = |y_{n1} - y_{n2}|$), if the effect of the within-family environment is uniform across pairs of twins of different genotypes. However, if pairs of twins of different genotypes react differently to similar environmental influences, then the expected value of the differences between MZ twin scores may depend on their genotype ($|y_{11} - y_{12}| \neq \ldots \neq |y_{n1} - y_{n2}|$). The sum of MZ twin scores, on the other hand, will differ across families if twins in different families have different genotypes, different (family) environments, or both. Jinks and Fulker (1970) reasoned that the absolute differences can be predicted from the sums, if interaction between genotype and (within-family) environment is present. By regressing the absolute differences on the sums, it is possible to test whether the relation between genes and environment is linear, that is, $(|y_{n1} - y_{n2}|) = b_0 + b_1(y_{n1} + y_{n2})$, or curvilinear, that is, $(|y_{n1} - y_{n2}|) = b_0 + b_1(y_{n1} + y_{n2}) + b_2(y_{n1} + y_{n2})^2$, where b_0 denotes the intercept, and b_1 and b_2 the regression parameters for the linear and curvilinear effect, respectively. A linear effect implies that the effect of the environment is stronger (i.e., gives rise to greater individual differences) at either higher or lower levels of the genotypic factor, depending on the sign of the regression parameter b_1. A pure curvilinear effect (i.e., no linear effect) implies that the effect of the environment is stronger at either intermediate levels or extreme levels of the genotypic factor, depending on the sign of the regression parameter b_2. Jinks and Fulker (1970) also noted that G × E can be detected through heterogeneity of within-twin standard deviations caused by means and standard deviations being related (p. 315). This means that when G × E is present, the variance in the difference scores will vary with the size of the sum scores. The scatter plots in Figure 1, where we plotted the absolute $twin_1$–$twin_2$ differences against the $twin_1$–$twin_2$ sums, illustrate the expected heterogeneity. When there is no G × E (i.e., no relation between the differences and the sums), the variance in the difference scores is more or less homogeneous across levels of the sum scores (Figure 1A). When linear G × E is present, the dispersion in the difference scores increases (Figure 1B) or decreases (Figure 1C) with increasing sum scores, depending on the direction of the effect. When curvilinear G × E is present, the dispersion of the difference scores is either larger for more extreme sum scores (i.e., small and large sum scores, Figure 1D), or for intermediate sum scores (Figure 1E).

Although the Jinks and Fulker test (henceforth JFT) is particularly easy to conduct, its application is hampered by low power, and sensitivity to nonnormality in the data (e.g., Boomsma & Martin, 2002; Martin, 1999; Purcell, 2002). For example, when the distribution of a trait is skewed due to, say, floor or ceiling effects of the test instrument, significant relations between the differences and the sums may appear that are not attributable to actual G × E interaction. Although this sensitivity to ('non-G × E') violations of normality is problematic, the question remains how well the JFT detects G × E when data are free of obvious sources of nonnormality, such as censoring.

The aim of the present article is to study the power of the JFT to detect G × E, and to compare these results to those obtained with an alternative test for detecting G × E, which is based on Marginal Maximum Likelihood (MML test, henceforth MMLT). First, we will discuss some of the assumptions which underlie the standard additive genetic twin models. Second, we will elaborate on the rationale of MML, and illustrate the implementation of MML in the univariate twin design. Third, we will study the power to detect G × E with the JFT and the proposed MMLT in simulated datasets, distinguishing between normally and nonnormally distributed data. Fourth, we will examine the influence of the normal scores transformation (an efficient normalizing transformation) on the detection of G × E with the MMLT and the JFT. We conclude the article with a brief discussion, in which we broach possible extensions of the MMLT.

Standard Additive Genetic Twin Model and Marginal Maximum Likelihood

In the standard additive genetic twin model, that is, the ACE-model,[1] where A, C, and E stand for additive (polygenic) genetic, and shared, and unshared environmental influences, respectively, any variance due to G × E will end up in the E-component of the model (e.g., Purcell, 2002). Intuitively, this seems correct because, as a function of the unshared environmental influence E, G × E can only contribute to differences between twins. In standardized twin variance component models, the additive genetic correlation $Cor(A_1,A_2)$ is fixed to 1 in MZ twins, and to .5 in dizygotic (DZ) twins, while the unshared environmental correlation $Cor(E_1,E_2)$ is 0 in both MZ and DZ twins. Following

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

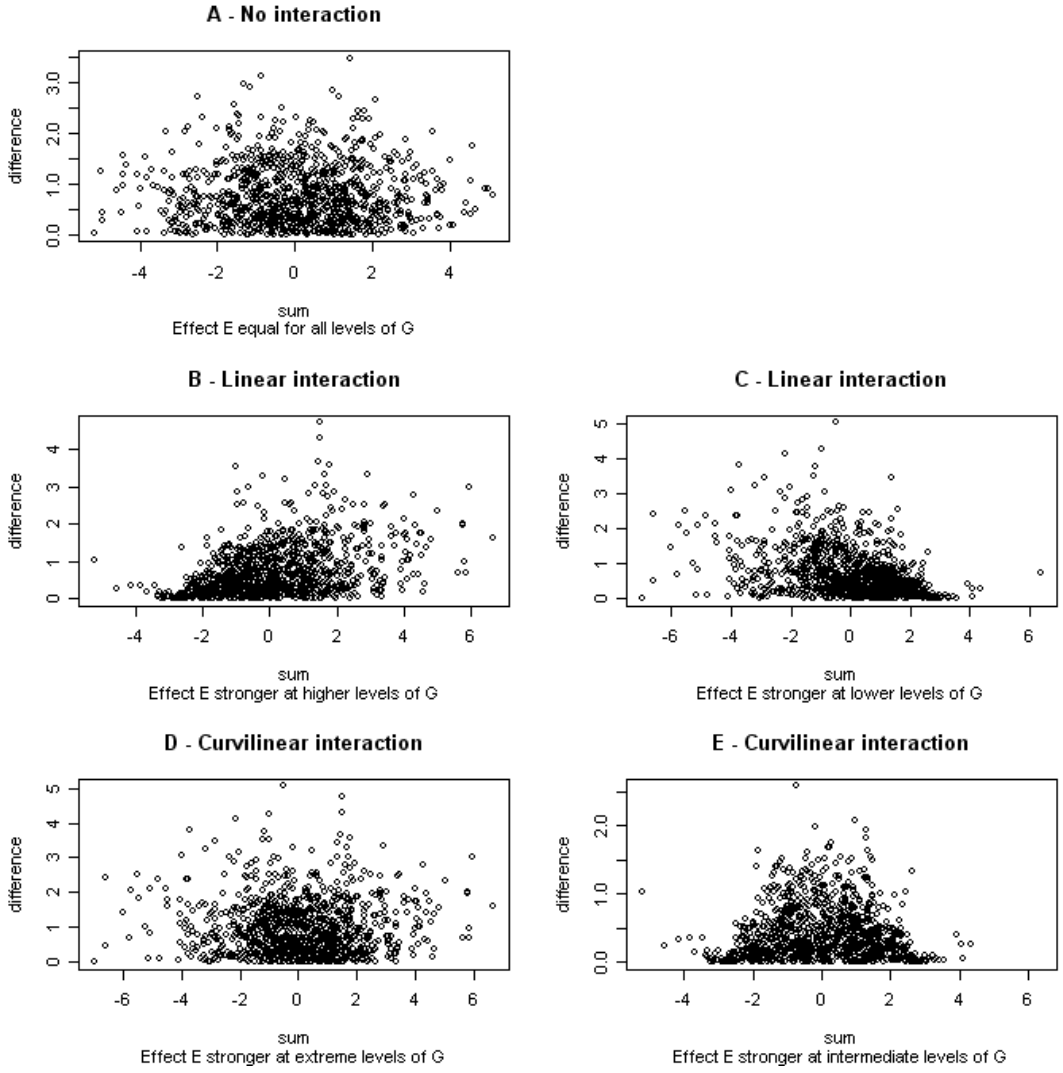

Figure 1

Plotting the absolute differences $|t_1 - t_2|$ against the sums $t_1 + t_2$, for N_{MZ} = 1000 cases, a^2 = .5, e^2 = .4 and 10% variance accounted for by G x E (that is, b^2 = .10).

Note: For these illustrations, c^2 was assumed to be zero.

the rules of covariance algebra (e.g., Kenny & Judd, 1986), it can be shown that $Cov(A_1E_1,A_2E_2) = Cov(A_1A_2)*Cov(E_1E_2) = 0$, that is, the relation between A and E in twin 1 and the relation between A and E in twin 2 are uncorrelated, and thus only attribute to differences within twin pairs. In the light of this, the key to detecting G × E will be in studying the E-component.

The JFT is based on MZ twins only. In the remainder, we assume that we only have data of MZ twins at our disposal. Under these circumstances, only a so-called AE-model (as depicted in Figure 2) can be fit, in which the genotypic factor A may include genetic as well as shared environmental effects, as those effects cannot be disentangled in MZ twins. To ease presentation, we assume that C is absent.

In normal theory maximum likelihood (ML) estimation of the parameters in the AE-model, the factors A and E, and thus the observed variables, are assumed to follow normal distributions. One implication of this normality assumption is that the residuals in the regression of the observed phenotype on the genotypic factor A are homoscedastic. This means that the proportion of variance explained by the unshared environment is independent of the level of the genotypic factor, that is, a single statistic (i.e., parameter *e* in Figure 2) is sufficient to describe the unique environmental effect for the whole population. The model parameters can be estimated maximizing the raw data log-likelihood function:

$$LogL = \sum_{i=1}^{N} \log f(y_i; \sigma_A^2, \sigma_E^2, \mu) = \sum_{i=1}^{N} \log\left[(2\pi)^{-1/2} \, |\Sigma|^{-1/2} \, e^{-1/2 (y_i-\mu)' \Sigma^{-1} (y_i-\mu)}\right] \quad [1]$$

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

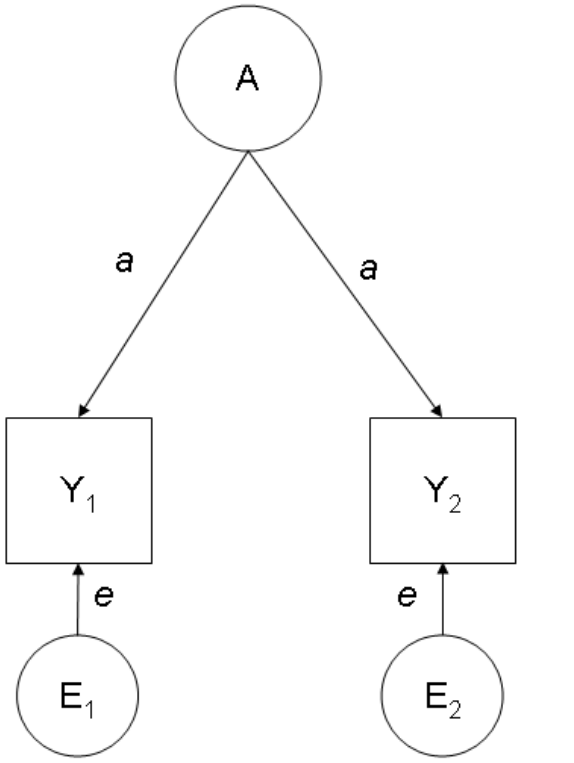

Figure 2

Path diagram of an AE-model for one pair of MZ twins.

Note: Because the genotypic factors of MZ twins are unity correlated, only 1 (i.e., collapsed) genotypic factor is shown in the diagram. In the presence of G × E interaction, the variance of the factors E will vary across levels of the genotypic factor A.

where $f(y_i;\sigma_A^2,\sigma_E^2,\mu)$ is the bivariate normal density function, σ_A^2 and σ_E^2 denote the additive genetic and unshared environmental variance respectively, μ the expected phenotypic mean vector, and Σ the expected covariance matrix. The variances in Σ equal $\sigma_A^2 + \sigma_E^2$, and the covariance equals σ_A^2. Generally, the phenotypic means are not modeled beyond the specification of a constraint to the effect that twin 1 mean equals the twin 2 mean. Assuming the absence of missing data and discarding the means (by prior centering, say), this log-likelihood function may also be expressed in terms of the phenotypic covariance matrix S, which is a sufficient statistic (Azzelini, 1996), that is, $(N\text{-}1)*[\log(|\Sigma|) + \text{trace}(\Sigma^{-1}S)$, where S is the observed MZ covariance matrix and N is the sample size.

The presence of G × E, however, as conceptualized here, implies that the amount of variance attributable to the unshared environment varies across levels of the genotypic factor. Therefore a single statistic is no longer sufficient to describe the unique environmental effect for the whole population, as the statistic (i.e., parameter e in Figure 2) assumes different values for different levels of the genotypic factor. The presence of G × E thus results in heteroscedasticity, as was illustrated in Figure 1. This in turn implies that the assumption of normality of the observed data (i.e., $\mathbf{y}_i \sim N[\mu,\Sigma]$) no longer holds. However, we can still assume normality conditional on the level of the genotypic factor, that is, $\mathbf{y}_i|\eta^* \sim N(\mu + I\otimes a^*\eta_i^*, \Sigma_E^*)$, where I is a 2 × 1 unit vector, η_i^* denotes a given level on the latent genotypic factor, a the genotypic factor loading (i.e., $a^2 = \sigma_A^2$), and Σ_E^* is the 2 × 2 diagonal covariance matrix for the residual variance, that is, $\Sigma_E^* = [\sigma_{\hat{E}}^{2*}\ \sigma_{\hat{E}}^{2*}]$, where $e^2 = \sigma_{\hat{E}}^{2*}$. Note that now the environmental variance $\sigma_{\hat{E}}^{2*}$ may be a function of the additive genetic factor η, that is, $\sigma_{\hat{E}}^{2*}$ can assume different values given different levels of the genotypic factor.

Given this assumption, one can use MML estimation (Bock & Lieberman, 1970), to estimate the parameters and model the heteroscedasticity. Although originally developed for item response theory modeling of dichotomous items, this method is equally applicable to continuous data (Hessen & Dolan, 2006). The bivariate normal distribution of the observed data, $f(y_i)$, is then expressed as the integral of the product of the conditional density of y_i given η_i and the density of η_i, that is,

$$f(y_i) = \int_{-\infty}^{\infty} f(y_i \mid \eta_i) f(\eta_i) d\eta \qquad [2]$$

Equation 2 may be recognized as an application of the law of multiplication (Miller & Miller, 2004). The density of η_i, $f(\eta_i)$ is the standard normal distribution, and $f(y_i|\eta_i)$ is the conditional distribution, that is, as explained above, $\mathbf{y}_i|\eta^* \sim N(\mu + I\otimes a^*\eta_i^*, \Sigma_E^*)$. Although this indefinite integral cannot be expressed in closed form, it can be evaluated to any practical degree of accuracy using Gauss-Hermite quadrature (Bock & Lieberman, 1970). This numerical method allows one to approximate the above integral by

$$f(y_i) \approx \sum_{j=1}^{J} f(y_i \mid \eta_i) f(\alpha_j) \approx \sum_{j=1}^{J} f(y_i \mid \alpha_j) w_j \qquad [3]$$

where α_j denotes the j^{th} Gauss-Hermite quadrature point, and w_j, that is, $f(\alpha_j)$, the corresponding weight. Note that in the present context, the Gauss-Hermite quadrature points may be interpreted as levels of the latent genotypic factor. Note also that the number of quadrature points j depends on the precision with which one wishes to approximate the integral. Various trials indicate that 10 quadrature points are sufficient to achieve a satisfactory approximation. The quadrature points and weights can be retrieved from tables (e.g., Abramowitz & Stegun, 1970), or from the Internet.[2] Previously, MML was used to study heteroscedasticity in the single common factor model (Hessen & Dolan, 2006). In this context, MML allowed the study of the precision of measurement of each indicator of the common factor conditional on the level of the common factor.

Equation 3 may be viewed as an approximation to the bivariate normal distribution $f(\mathbf{y}_i;\sigma_A^2,\sigma_E^2,\mu)$. This approximation may be viewed as a multivariate normal finite mixture distribution (e.g., Dolan

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

& van der Maas, 1998). Indeed, this is exactly the way this model is specified in Mx (Neale et al., 2003; see Appendix A). With this approximation in hand, we can define the following approximate log-likelihood function:

$$LogL \approx \sum_{i=1}^{N} \log \sum_{j=1}^{J} f(y_i \mid \alpha_j) w_j \qquad [4]$$

where w_j and $f(y_i|\alpha_j)$ are defined above. As discussed above, the parameters σ_A^2, σ_E^2, and μ enter into the function via $f(y_i|\alpha_j)$, that is, $y_i|\eta^* \sim N(\mu + I \otimes a^* \alpha_j, \Sigma_E^*)$. As in Equation 4, the parameters μ can be removed by prior centering.

We emphasize that our present aim is simply to *detect* G × E in terms of variation in the unshared environmental variance across the levels of the genotypic factor, and not necessarily to arrive at an exact description of this variation. However, in our present use of MML, we require some function to account for this change in variance. In the example Mx script (see Appendix A), and in the subsequent power analyses, we followed Hessen and Dolan (2006), and used an exponential function to model this variation in unshared environmental variance across the levels of the genotypic factor:

$$\sigma_E^{2*} = \exp(\phi_0^2 + \phi_1 * \alpha_j + \phi_2 * \alpha_j^2) \qquad [5]$$

where α_j again denotes the Gauss-Hermite quadrature points, ϕ_0 is the intercept, and ϕ_1 and ϕ_2 are the regression parameters for the linear and the curvilinear effect, respectively. One advantage of the exponential function is that it cannot assume negative values, which is convenient as we are estimating a variance term. Even though the exponential function may not be suitable for modeling exact linear change (although the function can be practically linear, depending on the size of ϕ_1 and ϕ_2, and the range of values of α_j for which one studies the function), it is useful for fitting monotonic increase or monotonic decrease in variance, and it can be used to describe parabolic change in variance. Figure 3 includes plots of the unshared environmental variance σ_E^2 against the levels of the genotypic factor. Other functions, given theoretical or practical considerations, may be specified instead of Equation 5.

In the MMLT of G × E, we first fit the model with ϕ_1 and ϕ_2 fixed to zero. This is a homoscedastic model as the environmental variance does not depend on the level of A, that is, $\sigma_E^{2*} = \exp(\phi_0)$. This model produces almost exactly the same results as the standard unconditional model (Equation 1). The factor that MML involves as an approximation (Equation 3) to the integral (Equation 2) introduces a slight discrepancy. Subsequently, we fit the model with ϕ_1, or ϕ_1 and ϕ_2, freely estimated, to accommodate the heteroscedasticity arising from G × E. As suggested by Hessen and Dolan (2006), a likelihood ratio (minus twice the difference in log-likelihood of the two models) can be used to test the statistical significance of the parameters ϕ_1 (an asymptotic $df = 1$ χ^2 test) and/or ϕ_1 and ϕ_2 (an asymptotic $df = 2$ χ^2 test).

Below we investigate the power to detect G × E interaction in MZ twin data using both the JFT and the method based on MML estimation.

The Power to Detect G × E: Design

The design of the simulation study comprises 4 × 3 × 3 conditions: four effect sizes of the interactive effect (0%, 2.5%, 5% and 7.5% of the total variance), three effect sizes of the additive polygenic genetic effect (20%, 50%, or 70% of the total variance explained), and three sample sizes ($N = 200$, $N = 400$, and $N = 800$, where N is the number of MZ pairs). In all conditions, the variance, which was not explained by additive genetic effects or G × E interaction, was attributed to unshared environmental effects. As mentioned, common environmental effects were not considered. Each condition was replicated 1000 times. Each simulated dataset was analyzed using four models: linear and nonlinear JFTs (based on F statistics), and linear and nonlinear MMLTs (based on χ^2 statistics). All analyses concerning the simulations were carried out using our own Fortran program, which included NPSOL (Gill et al., 1986) for optimization using exact gradients, and various IMSL (1991) routines for data simulation and analysis.

Data without G × E effects were simulated as follows:

$$\sigma_Y^2 = a^2 \times \sigma_A^2 + (1 - a^2) \times \sigma_E^2 \qquad [6]$$

where a^2 equals the heritability coefficient, $1 - a^2$ equals the unshared environmental effect e^2 (were a and e correspond to the parameters a and e in Figure 2), and both A and E were simulated as standard normal variates, that is, $N(0,1)$. Data without G × E interaction were used to estimate the probability of a false positive, given a nominal value of $\alpha = .05$.

Data with linear G × E interaction were simulated given the model:

$$\sigma_Y^2 = a^2 \times \sigma_A^2 + (1 - a^2 - b^2) \times \sigma_E^2 + b^2 \times \sigma_{A \times E}^2 \qquad [7]$$

where a^2 equals the heritability coefficient, and b^2 the standardized interactive effect. Note that *adding* $b^2 * \sigma_{A \times E}^2$ creates data in which the effect of the unshared environment is stronger for higher levels of the genetic factor, while *subtracting* $b^2 * \sigma_{A \times E}^2$ creates data in which the effect of the unshared environment is stronger for lower levels of the genetic factor.

Data with curvilinear G × E interaction were simulated following the model:

$$\sigma_Y^2 = a^2 \times \sigma_A^2 + (1 - a^2 - b^2) \times \sigma_E^2 + b^2 \times \sigma_{|A| \times E}^2 \qquad [8]$$

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

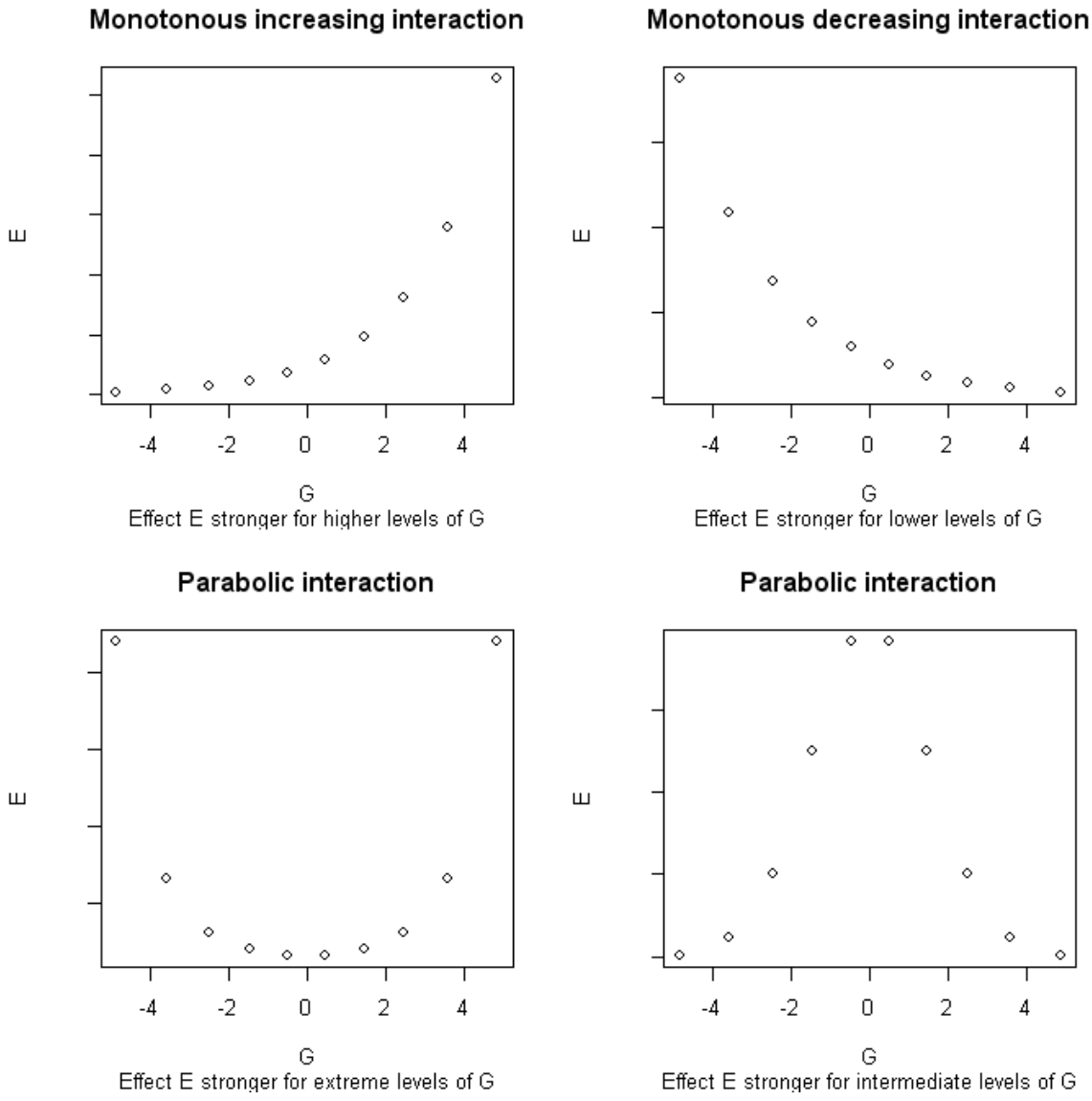

Figure 3

Plotting the conditional estimates of the unshared environmental variance for 10 levels of the genotypic factor for $N_{MZ} = 1000$ cases, $a^2 = .5$, $e^2 = .475$ and 2.5% variance accounted for by G × E (i.e., $b^2 = .025$).

Note: For these illustrations, c^2 was assumed to be zero.

where |A| in $\sigma^2_{|A|\times E}$, that is, the absolute values of A, gives rise to the curvilinear effect. Note that *adding* $b^{2*}\sigma^2_{|A|\times E}$ creates data in which the environmental effect is stronger for extreme levels of the genetic factor A, while *subtracting* $b^{2*}\sigma^2_{|A|\times E}$ creates data in which the environmental effect is stronger for intermediate levels of the genotypic factor A. Because with regard to the power to detect G × E the results only depend on the size of b^2 and not its sign, the results will be limited to simulations with additive G × E effects for both the linear and the curvilinear case.

Apart from analyzing these data with the JFT and MMLT, all simulated samples were tested for univariate normality using the Shapiro-Wilk test, which has been shown to be the most powerful test to detect nonnormality (Shapiro et al., 1968). Tests for univariate normality were conducted with a view to studying the sensitivity of both the JFT and the MMLT to deviations from normality.

Results

The results of the simulation study are shown in Table 1. For homoscedastic data, that is, data that include no G × E interaction, the JFTs and the MMLTs detect linear and curvilinear interaction in about 5% of the samples, independent of the sample size. Likewise, the Shapiro-Wilk test detects deviation from normality in 5% of the samples. Given a nominal α of .05, these results are in line with expectation.

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

Table 1

Power to Detect Genotype by Environment Interaction Using the Jinks and Fulker test (JFT), and Marginal Maximum Likelihood (MMLT)

		$N = 200$			$N = 400$			$N = 800$		
		Homosced.	Linear	Curvilinear	Homosced.	Linear	Curvilinear	Homosced.	Linear	Curvilinear
$a^2 = .20$										
$e^2 = .775$	*JFT lin*	.05	.45	.05	.05	.72	.06	.06	.97	.06
$b^2 = .025$	*JFT nonlin*	.06	.38	.10	.04	.62	.12	.05	.94	.21
	MMLT lin	.05	.86	.15	.05	.99	.14	.04	1	.16
	MMLT nonlin	.07	.80	.16	.05	.98	.25	.05	1	.44
	SW test	.05	.31	.03	.05	.46	.01	.04	.68	.02
$e^2 = .75$	*JFT lin*	.04	.74	.08	.05	.95	.07	.04	1	.09
$b^2 = .05$	*JFT nonlin*	.05	.65	.17	.07	.95	.20	.04	1	.34
	MMLT lin	.04	.98	.23	.05	1	.25	.04	1	.22
	MMLT nonlin	.05	.98	.27	.06	1	.44	.04	1	.69
	SW test	.05	.55	.03	.05	.72	.02	.05	.93	.01
$e^2 = .725$	*JFT lin*	.06	.83	.07	.06	.99	.08	.05	1	.08
$b^2 = .075$	*JFT nonlin*	.05	.80	.17	.05	.98	.27	.05	1	.43
	MMLT lin	.05	1	.35	.05	1	.37	.05	1	.38
	MMLT nonlin	.06	1	.38	.05	1	.59	.06	1	.87
	SW test	.05	.67	.04	.05	.87	.03	.06	.98	.01
$a^2 = .50$										
$e^2 = .475$	*JFT lin*	.04	.92	.07	.05	1	.09	.05	1	.08
$b^2 = .025$	*JFT nonlin*	.04	.88	.21	.04	1	.35	.05	1	.56
	MMLT lin	.04	.98	.09	.04	1	.11	.05	1	.13
	MMLT nonlin	.06	.99	.31	.06	1	.56	.06	1	.88
	SW test	.04	.57	.03	.05	.81	.03	.05	.96	.02
$e^2 = .45$	*JFT lin*	.05	1	.09	.04	1	.09	.05	1	.09
$b^2 = .05$	*JFT nonlin*	.06	.99	.34	.05	1	.51	.05	1	.77
	MMLT lin	.05	1	.13	.05	1	.16	.06	1	.14
	MMLT nonlin	.07	1	.50	.05	1	.82	.05	1	.98
	SW test	.05	.85	.06	.06	.97	.05	.06	1	.05
$e^2 = .425$	*JFT lin*	.04	1	.12	.05	1	.11	.05	1	.10
$b^2 = .075$	*JFT nonlin*	.04	1	.36	.05	1	.61	.05	1	.88
	MMLT lin	.06	1	.20	.04	1	.19	.06	1	.16
	MMLT nonlin	.07	1	.67	.04	1	.93	.06	1	1
	SW test	.07	.97	.10	.04	.99	.07	.06	1	.11
$a^2 = .70$										
$e^2 = .275$	*JFT lin*	.06	1	.10	.05	1	.10	.05	1	.11
$b^2 = .025$	*JFT nonlin*	.05	1	.43	.06	1	.66	.05	1	.88
	MMLT lin	.06	1	.13	.07	1	.12	.05	1	.13
	MMLT nonlin	.06	1	.49	.07	1	.78	.06	1	.98
	SW test	.06	.40	.05	.04	.58	.04	.05	.80	.03
$e^2 = .25$	*JFT lin*	.06	1	.11	.05	1	.11	.05	1	.12
$b^2 = .05$	*JFT nonlin*	.06	1	.55	.05	1	.81	.07	1	.97
	MMLT lin	.06	1	.16	.05	1	.14	.06	1	.17
	MMLT nonlin	.07	1	.72	.06	1	.96	.07	1	1
	SW test	.05	.61	.09	.06	.85	.11	.04	.97	.13
$e^2 = .225$	*JFT lin*	.04	1	.13	.05	1	.14	.05	1	.13
$b^2 = .075$	*JFT nonlin*	.06	1	.66	.05	1	.88	.05	1	.99
	MMLT lin	.05	1	.21	.05	1	.20	.05	1	.20
	MMLT nonlin	.07	1	.86	.06	1	.99	.06	1	1
	SW test	.05	.72	.14	.05	.93	.18	.05	.99	.31

Note: Expressed in percentage significant, based on 1000 replications.

N is the number of monozygotic twin pairs. a^2 denotes the percentage of variance explained by additive genetic effects, e^2 the percentage of variance explained by unique environmental effects, and b^2 the percentage of variance explained by the interaction between genes and environment.

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

To detect linearly modeled G × E interaction, the MMLT is more powerful than the JFT. This is most pronounced when sample size is small, and both additive genetic effects and G × E interaction explain only a small percentage of the total variance. For example, in samples consisting of 200 MZ twin pairs with additive genetic effects explaining 20% and G × E explaining only 2.5% of the total variance, MML detects this interaction in 86% of the cases, while the JFT detects the interaction in 45% of the cases. However, the power of both tests to detect linear G × E increases greatly when additive genetic effects are more substantial. For example, when additive genetic effects explain 50% of the total variance, the JFT and MML detect small G × E interaction effects (accounting for only 2.5% of the total variance), in 92% and 98% of the cases, respectively. Apparently, detecting a small G × E interaction effect from a large unshared environmental component is more difficult for both detection methods (but more so for the JFT).

Clearly, both the MMLT and the JFT have much less power to detect curvilinearly modeled G × E interaction. Especially when sample size is small, and both the additive genetic effect and the curvilinear interaction effect are small, both tests often fail to detect the interaction. For example, in samples consisting of 200 MZ twins with additive genetic factors explaining 20%, and the curvilinear interaction only 2.5% of the total variance, the JFT detects this interaction in 10%, and the MMLT in 16% of the cases. With sample sizes of N = 400 and N = 800, the JFT detects the interaction in 12% and 21% of the cases, respectively, while the MMLT detects the interaction in 25% and 44% of the cases, respectively. As with the linear interactive effect, the power of both tests to detect the interaction improves as the additive genetic effect increases and the unshared environmental component, from which the G × E is part, becomes comparatively smaller.

With regard to normality, results of the Shapiro-Wilk test show that linear G × E interaction results in a greater deviation from normality than curvilinear G × E. Note that the interaction effect is equal in terms of variance explained by the interaction effect. These results seem to indicate that the detection of G × E depends more on the resulting violation of normality than on the effect in terms of variance explained. To get some insight into the role of normality in the JFT and MMLT, the samples including linear G × E interaction were split up into the set in which the Shapiro-Wilk test was significant (given α = .05) and the set in which this test was not significant. The results obtained in these sets are shown in Table 2. Differences between the MMLT and the JFT in the power to detect G × E in normally distributed data are especially manifest when sample size is small (i.e., N = 200 or N = 400), and both the additive genetic effect and the interaction effect are small (i.e., a^2 = .20 and b^2 = .025, or .05); while the MMLT shows satisfactory power to detect linear G × E in normally distributed data, the power of the JFT is much lower, and often not sufficient. Both tests detect G × E more readily in nonnormally distributed data. However, while this difference is usually considerable for the JFT, it is mostly negligible for the MMLT as the power of this test seems overall satisfactory. When the additive genetic effect is large (70% of the total variance), both tests have power of 1 to detect G × E, irrespective of the normality of the data (not shown in Table 2).

Transformation: Normal Scores

As the simulation study above illustrates, data containing linear G × E interaction are inevitably not normally distributed. Deviations from normality are no rarity in psychological data and certainly not only associated with the presence of (G × E) interaction. Problems like censoring (i.e., truncation of the scale causing floor or ceiling effects), resulting in positive/negative skewness or kurtosis, are frequently encountered. Whatever the reasons for nonnormality of the scale distribution (e.g., poor test design, selective sampling, presence of interaction), violations of one of the most basic assumptions of all parametrical statistical tests (i.e., normally distributed residuals) may have consequences for the interpretability of results. For this reason, researchers often choose to transform their data to achieve normality. Although it is generally accepted that normalizing transformation may remove interactions (e.g., Martin, 1999), we would like to establish whether the effects of such a transformation on the power to detect G × E differ between the JFT and MMLT, as the results above suggest. To address this, we chose one of the most thorough procedures to normalize data, that is, the normal scores transformation, which is available in the LISREL program (Jöreskog & Sörbom, 2001). Figure 4 depicts the normalizing effect of the normal scores transformation. Because both the JFT and the MMLT demonstrated reasonable to perfect power to detect even small linear interaction effects in samples consisting of 400 MZ twins (see Table 1), we chose to limit the following simulations to samples of this size.

Table 3 shows the power results before and after the normal scores transformation. As we saw previously, the MMLT always detects linear G × E interaction when N = 400, even when the effect size is small. When additive genetic effects only explain 20% of the total variance, the JFT is somewhat less powerful, but for higher heritability coefficients, the JFT also detects G × E in all cases. As expected, the normalizing transformation greatly affects the results. When additive genetic effects explain only 20% of the total variance, neither the MMLT nor the JFT detect G × E interaction after transformation, regardless of the effect size of the interaction effect. When additive genetic effects explain 50% of the variance, the power of both the JFT and the MMLT to detect G × E improves somewhat, but is still far from satisfactory, especially for small interaction effects. However, when the additive genetic component becomes larger, and the unshared environmental component (of which the G × E effect is part) becomes comparatively smaller, the normalizing transformation no longer has a

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

Table 2

Results of Linear G × E as Presented in Table 1, Subdivided for Normally and Nonnormally Distributed Samples

		$N = 200$		$N = 400$		$N = 800$	
$a^2 = .20$		Normal $N_s = 689$	n-Normal $N_s = 311$	Normal $N_s = 541$	n-Normal $N_s = 459$	Norma $N_s = 316$	n-Normal $N_s = 684$
$e^2 = .775$	*JFT lin*	.35	.66	.62	.83	.91	.99
$b^2 = .025$	*JFT nonlin*	.28	.58	.51	.76	.85	.97
	MMLT lin	.81	.97	.99	1	1	1
	MMLT nonlin	.73	.95	.97	.99	1	1
		Normal $N_s = 449$	n-Normal $N_s = 551$	Normal $N_s = 284$	n-Normal $N_s = 716$	Normal $N_s = 73$	n-Normal $N_s = 927$
$e^2 = .75$	*JFT lin*	.59	.87	.89	.97	.99	1
$b^2 = .05$	*JFT nonlin*	.50	.78	.82	.94	.99	1
	MMLT lin	.97	.99	1	1	1	1
	MMLT nonlin	.95	1	1	1	1	1
		Normal $N_s = 326$	n-Normal $N_s = 674$	Normal $N_s = 130$	n-Normal $N_s = 870$	Normal $N_s = 25$	n-Normal $N_s = 975$
$e^2 = .725$	*JFT lin*	.71	.89	.96	.99	1	1
$b^2 = .075$	*JFT nonlin*	.67	.86	.95	.99	1	1
	MMLT lin	1	1	1	1	1	1
	MMLT nonlin	1	1	1	1	1	1
$a^2 = .50$		Normal $N_s = 428$	n-Normal $N_s = 572$	Normal $N_s = 186$	n-Normal $N_s = 814$	Normal $N_s = 38$	n-Normal $N_s = 962$
$e^2 = .475$	*JFT lin*	.90	.93	.99	1	1	1
$b^2 = .025$	*JFT nonlin*	.83	.91	.99	1	1	1
	MMLT lin	.97	.99	1	1	1	1
	MMLT nonlin	.98	.99	1	1	1	1
		Normal $N_s = 149$	n-Normal $N_s = 851$	Normal $N_s = 33$	n-Normal $N_s = 967$	Normal $N_s = 1$	n-Normal $N_s = 999$
$e^2 = .45$	*JFT lin*	1	.99	1	1	1	1
$b^2 = .05$	*JFT nonlin*	.98	.99	1	1	1	1
	MMLT lin	1	1	1	1	1	1
	MMLT nonlin	1	1	1	1	1	1
		Normal $N_s = 79$	n-Normal $N_s = 921$	Normal $N_s = 79$	n-Normal $N_s = 921$	Normal $N_s = 7$	n-Normal $N_s = 993$
$e^2 = .425$	*JFT lin*	1	1	1	1	1	1
$b^2 = .075$	*JFT nonlin*	1	1	1	1	1	1
	MMLT lin	1	1	1	1	1	1
	MMLT nonlin	1	1	1	1	1	1

Note: N is the number of monozygotic twin pairs. N_s is the number of simulated samples. JFT is Jinks and Fulker test, MMLT is test based on Marginal Maximum Likelihood. a^2 denotes the percentage of variance explained by additive genetic effects, e^2 the percentage of variance explained by unique environmental effects, and b^2 the percentage of variance explained by the interaction between genes and environment. For $a^2 = .7$, and $b^2 = .025/.05/.075$, both methods always detect G x E, independent of the normality of the data.

nullifying effect on the power to detect G × E, even when the interaction only explains a small part of the total variance. For almost all studied scenarios, the power of the JFT surpasses the power of the MMLT after the normal scores transformation, and while both tests show the tendency to detect nonlinear G × E after the normal scores transformation, while linear G × E was simulated, this tendency is most marked for the MMLT.

Discussion

The simulation studies showed that generally the power of the MMLT was greater than the power of the JFT in detecting linear and curvilinear G × E interaction, regardless of whether the Shapiro-Wilks test detected the inherent deviation from normality. Both the MMLT and the JFT detected G × E interaction more easily when the additive genetic effect was greater, and when the G × E interaction effect was linear. For our simulations, we defined effect sizes in terms of the percentage of variance explained. However, based on the present results, it is clear that the percentage of variance explained is not a good measure of effect size. The greater probability of detecting G × E in the circumstances mentioned is due to these circumstances resulting in a greater departure from normality. In the MMLT especially, it is the greater degree of heteroscedasticity that results in the greater probability of rejecting the hypothesis that either the linear or the

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

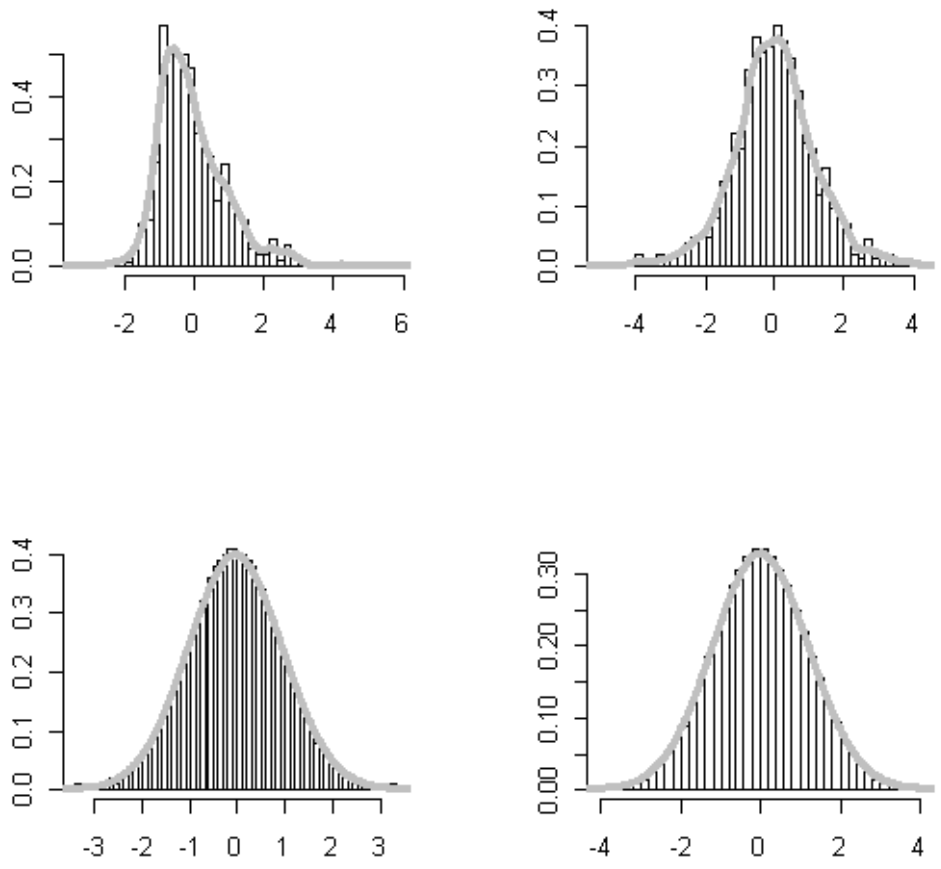

Figure 4

Illustrative histograms and density plot.

Note: Top left: N_{MZ} = 2000 cases, including linear G × E interaction (a^2 = .40, e^2 = .40, interaction accounts for 20% of the variance). The data display clear positive skew. Bottom left: histogram and density of same data following normal score transformation; the distribution no longer deviates from normality. Top right: 2000 cases, including curvilinear G × E interaction (a^2 = .50, e^2 = .40, interaction accounts for 10% of the variance). The data display kurtosis. Bottom right: histogram and density of same data following normal score transformation; the distribution no longer deviates from normality.

curvilinear interaction effect is nonsignificant ($\phi_1 = 0$ or $\phi_1 = \phi_2 = 0$, see Equation 4).

The effects of the normalizing transformation were generally in line with expectation: the transformation greatly reduced the power to detect the interaction. However, given a sufficiently large effect size, that is, a linear G × E in combination with relatively large genetic effects, we found that both the MMLT and the JFT continued to pick up G × E. The reason for this is that we applied univariate normal scores transformations. That is, the two observed scores were separately subjected to a normal scores transformation, as is the common practice. As is well known, however, perfect marginal normality does not guarantee multivariate, or in the present case, bivariate normality. Explorations of the data indicated that especially the MZ twin difference scores were not normally distributed (as it would be in case of bivariate normality). This is illustrated for a single simulated dataset in Figure 5, where the distribution of the MZ differences in normal scores are clearly leptokurtic (kurtosis = 2.45). The reason that the JFT performs slightly better in terms of power after the normal scores transformation is presumably because it focuses on these differences scores, whereas the MMLT is based on the full bivariate distribution.

For the present simulations, data were simulated according to an AE-model. Given MZ but not DZ twin data, modeling options are limited to the AE-model, that is, a restrictive, but not uncommon model. An important question is what influence the presence of shared environmental effects (C), or dominance effects (D) could have on the power to detect G × E in MZ twins. Since both the common environment and the dominance effects are fully shared in MZ twins, these effects will end up in the 'A'-factor of the AE-model. This factor is then a convolution of effects: additive genetic, common environmental, dominance and other nonadditive genetic factors. Our simulations showed that the power to detect an interaction improved when the additive genetic factor became comparatively large. However, both with the present MMLT and the JFT, one will not be able to distinguish between A × E, C × E, or D × E, because the A, C, and D effects are combined in one factor, and thus

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

Table 3

Power to Detect Linear G × E Before and After Normal Scores Transformation for Different Sizes of the Additive Genetic Effect (a^2), the G × E Interactions (b^2), and the Unshared Environment ($e^2 = 1 - a^2 - b^2$)

	$b^2 = .025$		$b^2 = .05$		$b^2 = .075$	
	Raw	Trans	Raw	Trans	Raw	Trans
$a^2 = .20$						
JFT lin	.76	.03	.95	.05	.99	.05
JFT nonlin	.66	.01	.91	.02	.98	.02
MMLT lin	.99	.00	1	.00	1	.01
MMLT nonlin	.98	.02	1	.13	1	.37
SW test	.49	.00	.74	.00	.86	.00
$a^2 = .5$						
JFT lin	1	.13	1	.37	1	.63
JFT nonlin	1	.11	1	.41	1	.77
MMLT lin	1	.02	1	.11	1	.29
MMLT nonlin	1	.02	1	.25	1	.71
SW test	.78	.00	.97	.00	.99	.00
$a^2 = .6$						
JFT lin	1	.48	1	.90	1	.99
JFT nonlin	1	.47	1	.93	1	.99
MMLT lin	1	.34	1	.80	1	.95
MMLT nonlin	1	.39	1	.92	1	.94
SW test	.73	.00	.94	.00	.99	.00
$a^2 = .7$						
JFT lin	1	.92	1	1	1	1
JFT nonlin	1	.92	1	1	1	1
MMLT lin	1	.89	1	1	1	1
MMLT nonlin	1	.90	1	1	1	1
SW test	.59	.00	.81	.00	.91	.00

Note: Sample size is $N = 400$ for all simulated samples.

N is the number of monozygotic twin pairs. JFT is Jinks and Fulker test, MMLT is test based on Marginal Maximum Likelihood. a^2 denotes the percentage of variance explained by additive genetic effects, e^2 the percentage of variance explained by unique environmental effects, and b^2 the percentage of variance explained by the interaction between genes and environment.

confounded. The presence of C or D in the observed data will thus complicate the interpretation of the results. A partial solution to this problem may lie in the extensions of the MMLT and JFT with additional data (e.g., DZ twin or adoption data).

The issue remains whether G × E detection methods are useful, given their sensitivity to (normalizing) scale transformations. Our position is that these methods are useful in the situation that there is no obvious source of nonnormality, such as censoring or poor scaling.[3] For instance, if one can demonstrate that G × E, as defined in the JFT or MMLT, is absent, one can certainly have a little more confidence in validity of the results of genetic covariance modeling. A significant JFT or MMLT result, even though it may be due to a variety of factors, is a cause for concern regardless of its exact cause. Normalizing transformations may remove this effect, but they will also affect all other results, for example, increasing or decreasing a^2. This exact effect of the transformation on all other results is likely to differ depending on the exact transformation (e.g., Box-Cox vs. normal score transformation), as different transformations may result in hard to interpret variation in a^2. There may of course be situations in which the choice for a nonlinear transformation of the data is based on substantive considerations, rather than on distributional concerns. For example, logarithmic transformation are suitable if one wishes to study proportional change rather than absolute change[4] (Falconer & Mackay, 1996; Zar, 1999) or if one wishes to study genes that act in a multiplicative fashion with a simple additive genetic model[5] (Mather & Jinks, 1977), while a square root transformation may be suitable if the character under study is effectively an area rather than a character of linear dimensions[6] (Mather & Jinks, 1977). Yet, the fact that nonnormality of the data may either *indicate* the presence of G × E (i.e., G × E being the source of the nonnormality), or *mimic* the presence of G × E

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

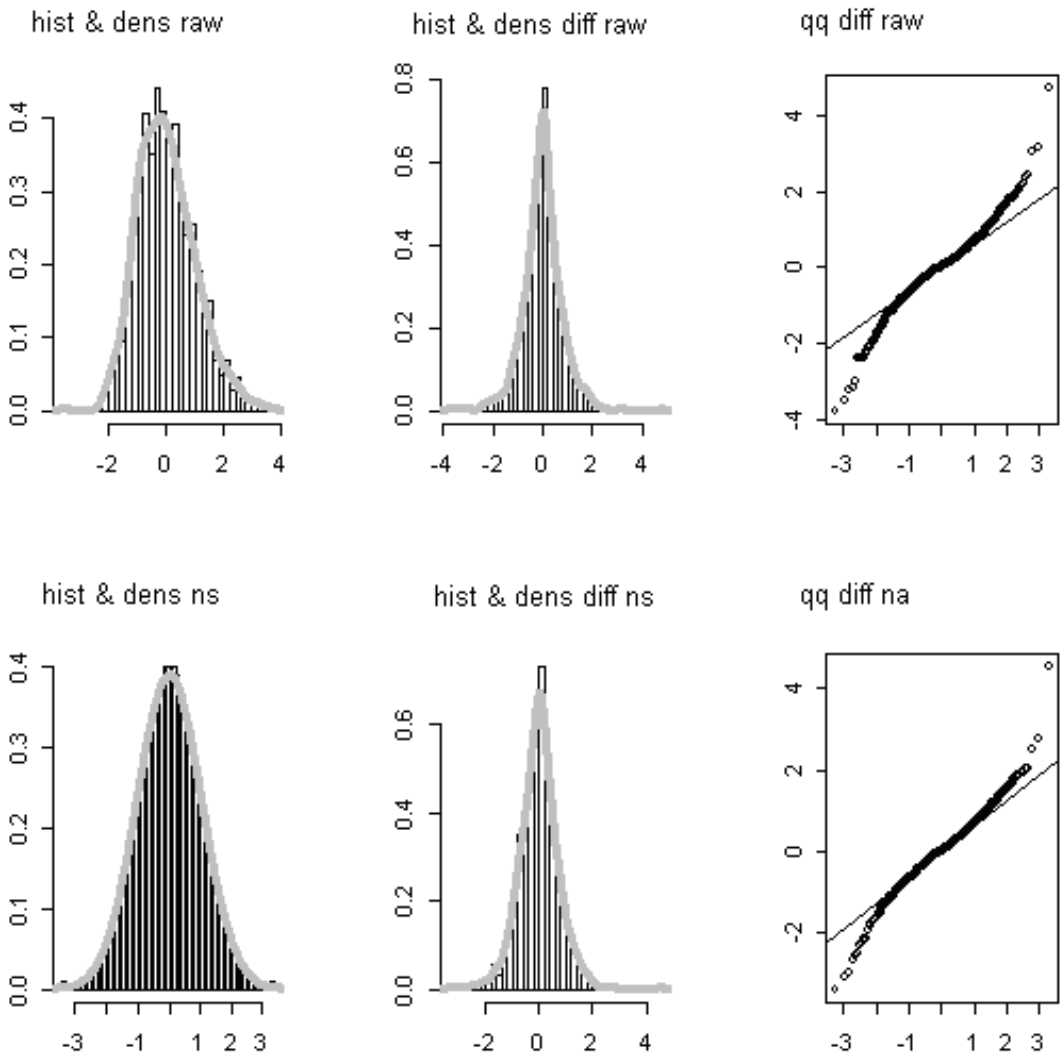

Figure 5

Illustration of the effect of the normal scores transformation with $N_{MZ} = 1000$, $a^2 = .70$, $e^2 = .25$, and 5% of the variance accounted for by G × E (i.e., $b^2 = .05$).

Note: Top row, left to right: Marginal raw distribution of twin 1, raw MZ difference score, QQ-plot of the raw MZ difference score. Bottom row: Marginal normal score distribution of twin 1, MZ differences in normal scores, QQ-plot of the differences in normal scores. The normal scores transformation renders the marginal distribution (bottom left) perfectly normal. However, the normal score differences remain highly leptokurtic (kurtosis equals 2.45; kurtosis of the raw differences equals 3.13). The QQ-plot (bottom right) also shows the peakedness of the distribution.

(i.e., G × E detection tests pick up a significant G × E effect, but factors other than G × E are the actual source of the nonnormality), remains worrisome. For now, the use of psychometrically sound measurement instruments would appear to be one way to minimize the problem of scale-dependency (van den Berg et al., 2006).

A second important issue, discussed by Eaves (1984) in the light of plant studies, is the possibility that the genes that control average performance may very well differ from the genes that control the sensitivity to the environment (i.e., the genes giving rise to the heteroscedasticity, see Berg et al., 1989, for a similar distinction between 'level' and 'variability' genes). Eaves (1984) noted that the JFT is limited in the sense that it will only reveal interaction effects if the same set of genes controls *both* (mean level and sensitivity). This limitation applies equally to MMLT as presented here.

Like the JFT, the MMLT is quite easy to carry out in the Mx program (Neale et al., 2003). Its main advantage when applied to MZ twin data is its greater power, when applied to untransformed data. In addition, the MMLT is amenable to various extensions. These include multigroup analysis, multivariate data, and the addition of DZ twins, which will allow one to include a component for the shared or common environment (i.e., C). In addition, by viewing the quadrature point as average levels due to one set of genes, it may be possible (with DZ twin data) to investigate whether the residual variance, in the present article denoted as purely environmental ($\sigma^{2}_{E}{}^{*}$), includes a genetic component. If so, this would suggest that the

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

genes controlling the average level are indeed distinct from those controlling the sensitivity to the environment. If the residual variance does, however, not include an additional genetic component, this would suggest that the finding of different genes controlling average performance and sensitivity to the environment, while well established in plant research, does not hold in outbred populations.

Endnotes

1 Dominance variance may also be modeled, albeit if only data on MZ and DZ twins are available, not simultaneously with C, for reasons of identification.

2 Note that the nodes and weights are usually given for the error function, not the standard normal distribution function. But, as Bock and Lieberman (1970) explain, a simple transformation, render these applicable for the standard normal distribution.

3 If in an average sample, depression, say, is measured with a clinical depression scale that contains many extreme items like 'I often feel that life is not worth living' or 'My future seems hopeless', the majority of respondents will respond negatively, and the distribution of the scores will be skewed for this sample. Censoring, that is, truncation of the score distribution, will also result if one uses a test that is much too difficult or too easy for the sample under study. Finally, skewness or kurtosis may result if one construes items with limited answer options ('never, sometimes, always'; 'yes, no'), as these options may not describe the respondents attitude accurately, and some options may be much more popular than others. In all examples, an obvious source of nonnormality would be the choice of instrument or scale rather than anything else.

4 Consider a reading intervention study, with dyslexic children reading 10 words before, and 15 words after the intervention, and nondyslexic children reading 30 words before, and 45 words after the intervention. Studying the absolute change in number of words read correctly would lead one to conclude that nondyslexic children profited more from the intervention than dyslexic children. However, studying the *proportional change* in reading speed, one would conclude that both groups profited equally well, as the speed of both groups increased by 50%.

5 Consider the following taken from Mather and Jinks (1977). If two genes act in a multiplicative fashion, and their joint effect is the product ($x_a x_b$) of their individual actions rather than the sum ($x_a + x_b$), then a simple additive genetic model is not appropriate. However, if we replace the measurement of the phenotype with its logarithm, that is, $\log(x_a x_b) = \log(x_a) + \log(x_b)$, then the multiplicative action has been removed, the genes now make their individual (independent) contribution, and a simple additive model will be suitable.

6 Consider the following taken from Mather and Jinks (1977). If genes are additive in their effect on the linear dimension of a phenotype (e.g., wing size), while the character that we study is effectively an area (i.e., the surface of the wing), the phenotype will reflect not the sum of the genes effect (as a linear character would), but the square of the sum. Replacing the observed phenotype with its square root would restore its linear basis, that is, the additive action of the genes. The simple additive genetic model would then fit the rescaled results.

Acknowledgments

Preparation of this manuscript was financially supported by NWO/MaGW VIDI-016-065-318 and NIH grants MH-65322 and DA-18673.

References

Abramowitz, M., & Stegun, I. (1970). *Handbook of mathematical functions*. New York: Dover Publications.

Azzelini, A. (1996). *Statistical inference based on the likelihood*. London: Chapman and Hall.

Berg, K., Kondo, I., Drayna, D., & Lawn, R. (1989). 'Variability gene' effect of cholesteryl ester transfer protein (CEPT) genes. *Clinical Genetics, 35*, 437–445.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika, 35*, 179–197.

Boomsma, D. I., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics, 3*, 872–882.

Boomsma, D. I., de Geus, E. J. C., van Baal, G. C. M., & Koopmans, J. M. (1999). A religious upbringing reduces the influence of genetic factors on disinhibition: Evidence for interaction between genotype and environment on personality. *Twin Research, 2*, 115–125.

Boomsma, D. I., & Martin, N. G. (2002). Gene-environment interactions. In H. D'haenen, J. A. den Boer, & P. Willner (Eds.), *Biological Psychiatry* (pp. 181–187). London: John Wiley & Sons.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-htt gene. *Science, 301*, 386–389.

Dolan, C. V., & van der Maas, H. J. L. (1998). Fitting multivariate normal mixtures subject to structural equation modelling. *Psychometrika*, *63*, 227–253.

Eaves, L. J. (1984). The resolution of genotype × environment interaction in segregation analysis of nuclear families. *Genetic Epidemiology, 1,* 215–228.

Eaves, L. J., Eysenck, H. J., & Martin, N. G. (1989). *Genes, culture & personality: An empirical approach*. London: Academic Press.

Falconer, D. S., & Mackay, T. F. C. (1996*). Introduction to quantitative genetics* (4th ed.). London: Longman Group.

Gill, P. E., Murray, W., Saunders, M. A., & Wright, M. H. (1986). *User's guide for NPSOL* (Version 5.0-2;

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

Technical report SOL 86-2). Stanford, CA: Department of Operations Research, Stanford University.

Heath, A. C., Eaves, L. J., & Martin, N. G. (1998). Interaction of marital status and genetic risk for symptoms of depression. *Twin Research, 1*, 119–122.

Hessen, D., & Dolan, C. V. (2006). *A heteroscedastic one-factor model and marginal maximum likelihood estimation of the parameters*. Manuscript submitted for publication.

IMSL. (1991*). IMSL STAT/LIBRARY Users manual* (Version 2). Houston: Author.

Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, mava, and classical approaches to the analysis of human behavior. *Psychological Bulletin, 73*, 311–349.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8.50: User's reference guide*. Chicago: Scientific Software International.

Kendler, K. S., Aggen, S. H., Jacobson, K. C., & Neale, M. C. (2003). Does the level of family dysfunction moderate the impact of genetic factors on the personality trait of neuroticism? *Psychological Medicine, 33*, 817–825.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99*, 422–431.

Martin, N. (1999). Gene-environment interaction and twin studies. In T. D. Spector, H. Snieder, & A. J. MacGregor (Eds.), *Advances in twin and sib-pair analysis*. London: Greenwich Medical Media.

Mather, K., & Jinks, J. L. (1977). *Introduction to biometrical genetics*. London: Chapman and Hall.

Miller, I., & Miller, M. (2004). *John E. Freund's mathematical statistics with applications* (7th ed.). NY: Prentice Hall.

Molenaar, P. C. M., & Boomsma, D. I. (1987). Application of nonlinear factor-analysis to genotype environment interaction. *Behavior Genetics, 17*, 71–80.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht, NL: Kluwer Academic Publishers.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5*, 554–571.

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A comparative study of various tests of normality. *Journal of the American Statistical Association*, *63*, 1343–1372.

Silberg, J. L., Rutter, M., Neale, M. C., & Eaves, L. J. (2001). Genetic moderation of environmental risk for depression and anxiety in adolescent girls. *British Journal of Psychiatry, 179*, 116–121.

Turkheimer, E., Haley, A., Waldorn, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*, 623–628.

Van den Berg, S. M., Glas, C. A. W., & Boomsma, D. I. (2006). *Variance decomposition using an IRT measurement model*. Manuscript submitted for publication.

Yaffe, K., Haan, M., Byers, A., Tangen, C., & Kuller, L. (2000). Estrogen, apoe, and cognitive decline: Evidence for gene-environment interaction. *Neurology, 54*, 1949–1954.

Zar, J. H. (1999). *Biostatistical analysis* (4th ed.). London: Prentice Hall.

https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

Appendix A

The Mx script, presented here, is an implementation of the MML method to detect G × E interaction in MZ twin data, given an AE model. The quadrature points and weights were obtained from the site http://www.efunda.com/math/num_integration/findgausshermite.cfm. However, bear in mind that these are provided for the error function, not the standard normal distribution. The transformation to values suitable for the standard normal distribution, that is, the values used below, is simple (see Bock & Lieberman, 1970). As mentioned, the approximate density function defined in Equation 3, may be viewed as a finite mixture distribution. The present Mx script defines the model in terms of a finite mixture. We provide comments to underline this.

```
#define nfac 1    ! number of factors equals 1, i.e., the additive genetic factor
#define nv 2      ! number of variables in model, i.e., mztwin1 and mztwin 2
#define nq 10     ! number of quadrature points - may be viewed as the number of
                  ! components in the mixture
#ngroups 2        ! number of groups in run

!-------------------------------------------------------------------------------

G1: Calculate stacked cov. and means
Calculation
Begin Matrices;
   J iden nv nv fi          ! auxiliary idenity matrix
   I unit nq 1 fi           ! auxiliary matrix
   A full nq 1 fi           ! Abscissae quadrature points
   L full nv nfac  fr       ! additive genetic factor loadings
   T full nv 1 fr           ! observed mean
   B full 1 1 fi            ! latent factor mean, fixed to zero

   P Stan nfac nfac         ! variance of additive genetic factor = 1

   F diag nv nv fr          ! unique environmental variance (equal for all levels of
                            ! factor A; parameter 'b0' as described in the
                            ! introduction)
   G diag nv nv fr          ! linear GxE regression parameter
                            ! (parameter 'b1' as described in the introduction)
   H diag nv nv fr          ! curivelinear GxE regression parameter
                            ! (parameter 'b2' as described in the introduction)
End Matrices;

! Specify quadrature points for MML
! i.e., points on the standard normal distribution

Matrix A
   -4.859462803   -3.581823472   -2.484325815   -1.465989067   -0.484935687
    0.484935708   1.465989094   2.484325842   3.581823484   4.859462828

! fix mean genotypic factor to zero
ma B
0

! equate parameters for twin1 & twin2
Equate  L 1 1 L 2 1
Equate  F 1 1 F 2 2
Equate  G 1 1 G 2 2
Equate  H 1 1 H 2 2
Equate  T 1 1 T 2 1

! provide starting values for parameters
Start .5 L 1 1 L 2 1
Start .8 F 1 1 F 2 2
```


https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press

```
Start .1 G 1 1 G 2 2
Start .6 H 1 1 H 2 2
Start 0 T 1 1 T 2 1

Begin Algebra;
   M = L'@(A*P)+L'@(I*B)+T'@I;                    ! Means
   S = \exp(I@F+A@G+(A.A@H)).(I@J) ;               ! Variance
End Algebra;
END

!-------------------------------------------------

G2: data group
Data Ninput=2  Nmodel= nq   ! Nmodel may be viewed as number of components
Rec File=MZtwins.dat
! N.B. MZtwins.dat is an ascii file with records of the form
! 1.8 2.7
! i.e. twin1 and twin2 on the same line
LAbels
   twin1 twin2
Begin Matrices;
   S comp = S1
   M COMP = M1
   W FULL nq 1
End Matrices;

! Specify Gauss-Hermite weights, interpretable as mixing proportions

Matrix W
      4.3107E-06 7.5807E-04 1.9112E-02 1.3548E-01 3.4464E-01
      3.4464E-01 1.3548E-01 1.9112E-02 7.5807E-04 4.3107E-06

Mean M;         ! means vectors of 10 components stacked
Covariance S;   ! cov. matrices of 10 components stacked
Weight W ;      ! weights interpretable as mixing proportions

option multiple issat
End;

! various tests…

save MML.mxs

get MML.mxs
drop H 1 1 1
end

get MML.mxs
drop G 1 1 1
end
```


https://doi.org/10.1375/twin.9.3.377 Published online by Cambridge University Press