

Genotype × Environment Interaction in Psychiatric Genetics: Deep Truth or Thin Ice?

Lindon Eaves

Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, VA, USA

Background: There continues to be significant investment in the detection of genotype × environment interaction ($G \times E$) in psychiatric genetics. The implications of the method of assessment for the genetic analysis of psychiatric disorders are examined for simulated twin data on symptom scores and environmental covariates. **Methods:** Additive and independent genetic and environmental risks were simulated for 10,000 monozygotic (MZ) and 10,000 dizygotic (DZ) twin pairs and the 'subjects' administered typical simulated checklists of clinical symptoms and environmental factors. A variety of standard tests for $G \times E$ were applied to the simulated additive risk scores, sum scores derived from the checklists and transformed sum scores. **Results:** All analyses revealed no evidence for $G \times E$ for latent risk but marked evidence for $G \times E$ and other effects of modulation in the sum scores. These effects were all removed by transformation. An integrated genetic and psychometric model, accounting for both the causes of latent liability and a theory of measurement, was fitted to a sample of the simulated sum-score data and showed that there was no significant modulation of the parameters of the genetic model by environmental covariates (i.e., no $G \times E$). **Conclusions:** Claims to detect $G \times E$ based on analytical methods that ignore the theory of measurement must be subjected to greater scrutiny prior to publication.

■ **Keywords:** psychiatric genetics, $G \times E$ interaction, twins, psychometrics

The last two decades have witnessed many research initiatives and claims to demonstrate the importance of genotype × environment interaction ($G \times E$) for human behavior and its disorders. The intimate connection between statistical demonstrations of $G \times E$ and the instruments employed to measure behavior has received less attention. This ambiguity has long been recognized by plant and animal geneticists (see, e.g., Mather & Jinks, 1982), but its implications for the substantive significance of claims to detect $G \times E$ for human behavior have been virtually ignored.

With the aid of simulated data, this article shows that the problem of inferring $G \times E$ is acute for the types of data (symptom counts and checklists) that form the mainstay of most psychiatric genetic epidemiology, to the point that claims to find $G \times E$ for psychiatric outcomes must seriously be questioned in the absence of any deeper analysis of the underlying biology or the relationship between behavioral dimensions and the instruments used to assess them.

Outline of Approach

(1) Simulated twin data are generated for a continuous latent behavioral trait and a continuous correlated index

of environmental risk on the assumption that a measured environmental covariate does not modulate any contributions of genes, shared family environment, or individual specific environment to differences in the simulated behavioral trait. (2) The latent trait and environmental index are 'scored' using a simple test procedure simulating those frequently used in behavioral measurement to generate symptom counts and environmental checklist scores with properties similar to those often found in psychiatric genetic epidemiology. (3) A series of statistical tests and analyses are performed on the simulated liabilities, derived test scores and transformed test scores to detect and estimate the main effects of genes, measured and unmeasured environment and the apparent modulation ('interaction') of genetic and environmental contributions by the environmental covariate. (4) An integrated model is developed and tested that recovers both the parameters of the genetic

RECEIVED 18 February 2017; ACCEPTED 20 February 2017

ADDRESS FOR CORRESPONDENCE: Lindon Eaves, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, PO Box 980003, Richmond VA 23298, USA. E-mail: eaves.lindon@gmail.com

TABLE 1
Expected Covariance Matrices for Simulated Environmental Risk and Liability to Behavioral Outcome in MZ and DZ Twins

		Monozygotic pairs				Dizygotic pairs			
		Environment (E)		Outcome (P)		Environment		Outcome	
		Twin 1	Twin 2	Twin 1	Twin 2	Twin 1	Twin 2	Twin 1	Twin 2
E	Twin 1	1.0000	0.7500	0.5000	0.3750	1.0000	0.7500	0.5000	0.3750
	Twin 2	0.7500	1.0000	0.3750	0.5000	0.7500	1.0000	0.3750	0.5000
P	Twin 1	0.5000	0.3750	1.0000	0.6875	0.5000	0.3750	1.0000	0.4375
	Twin 2	0.3750	0.5000	0.6875	1.0000	0.3750	0.5000	0.4375	1.0000

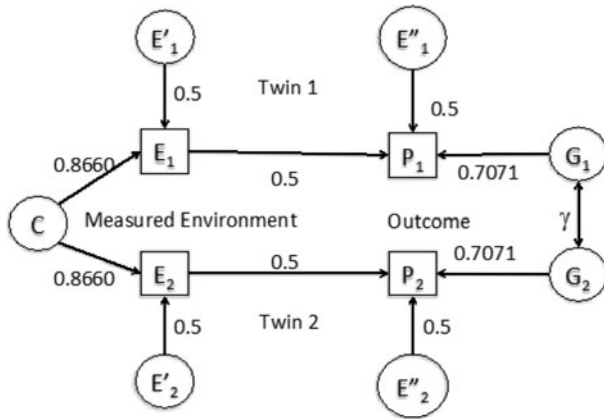


FIGURE 1
 Linear additive model for effects of genes, measured, and residual environment used to generate simulated twin data. Note: Subscripts 1 and 2 refer to first and second twins. Measured variables (represented by squares) are the measured environment (E) and the outcome phenotype (P). Latent random variables are additive genetic effects (G), shared environmental effects (C), and residual, unmeasured, environmental influences (E' and E'') on the measured environment and outcomes respectively. The standardized path coefficients are those used to simulate the data (see Table 1).

model, allowing for any 'true' $G \times E$ on the underlying liability, and the parameters of a typical measurement model relating the responses of subjects to survey items to differences in the underlying variables they are designed to assess.

Data Simulation

Simulated pairwise data vectors were generated for environmental covariates and continuous outcomes for each of 10,000 monozygotic (MZ) and 10,000 dizygotic (DZ) twin pairs from multivariate normal distributions with zero mean vectors and covariance matrices shown in Table 1. The underlying 'true' model assumes that there is a main effect of the covariate on outcome but no modulation of the contributions of genes and environment by the covariate. The expected covariance matrices correspond to a path model (Figure 1) in which the environmental covariate is

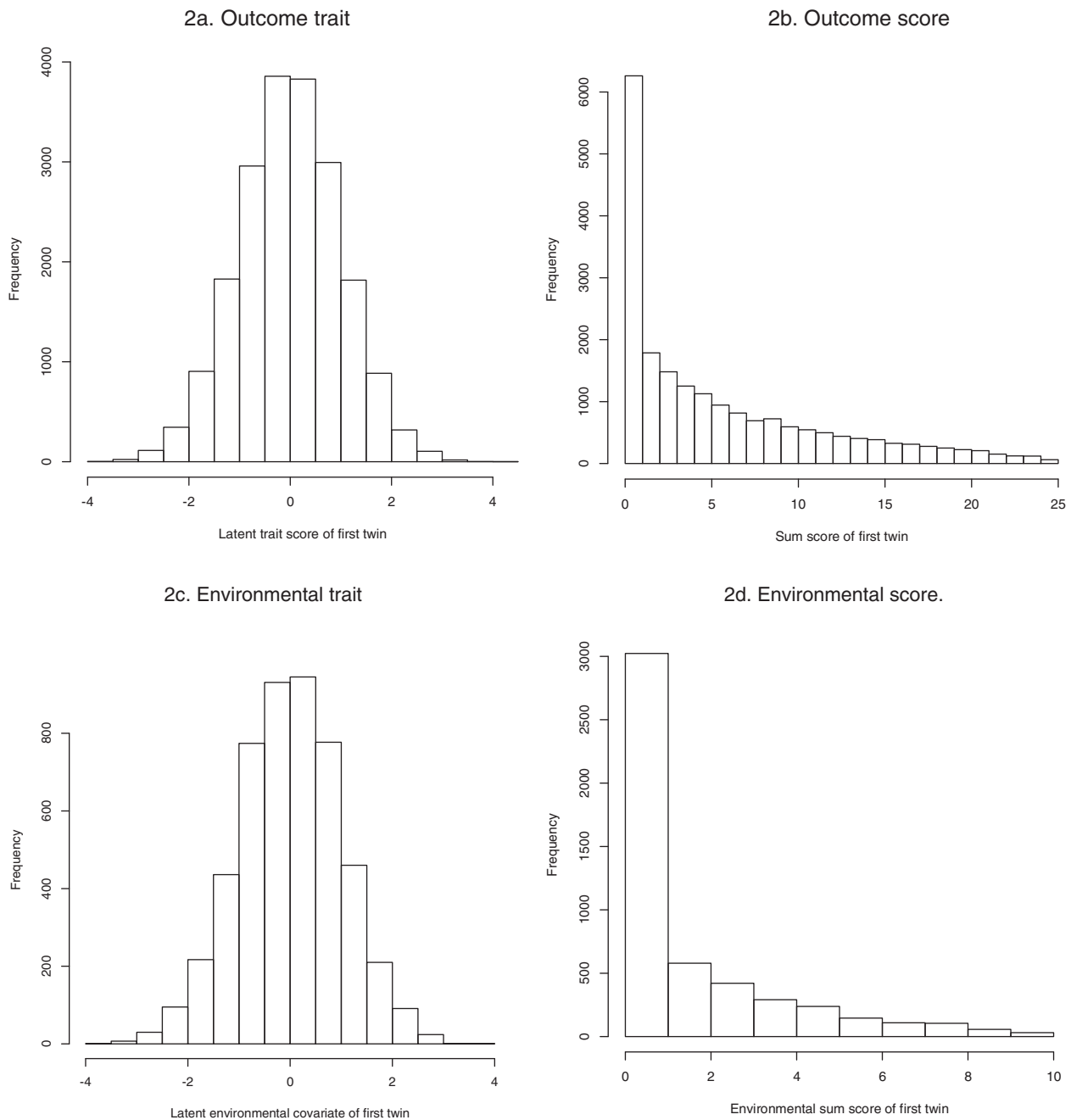
equally correlated between MZ and DZ twins (i.e., there is no within-person, genotype–environment correlation for the covariate) and explains all of the shared environmental component in the outcome. The model allows genetic and residual environmental effects on the outcome but no modulation of the genetic and environmental path coefficients by the environmental covariate (i.e., no $G \times E$ interaction or other modulation).

In practice, researchers do not have direct access to the 'true' dimensions of liability but attempt to measure them with plausible 'instruments'. In psychiatric genetic epidemiology, the instruments typically comprise checklists of symptoms and/or risk factors such as life events or features of the social environment. For the purposes of this study, we assume that the 'environment' is assessed by a checklist of 10 dichotomous items and that behavior is assessed by a list of 25 dichotomous symptoms. These numbers were chosen to correspond to some in one of our own studies. Responses to multi-item checklists of covariates and symptoms were simulated for each twin on the assumption that the probability of a subject endorsing the k th item is a monotonic function of increasing subject liability, θ . In our case θ corresponds to the subject's simulated 'true' covariate or outcome value. The probability that the i th subject endorses the k th item is thus assumed to be:

$$P_{ik} = \Phi[s_k(\theta_i - b_k)] \text{ (c.f. Lord \& Novick, 1968),}$$

where $\Phi(\theta)$ is the standardized cumulative normal p.d.f. (normal ogive), b_k is the value of θ at which $\Phi(\theta)$ changes most rapidly (also referred to as the 'item difficulty' or 'threshold') and s_k is the rate of change of Φ at b_k (also referred to as the 'sensitivity' or 'discriminating power' of the item).

The simulation assumed that the item parameters were the same for all items in a checklist. Thus, for the environmental checklist, it was assumed that $b_k = 1.3$ and $s_k = 1$ for all k . Similarly, all the items in the simulated behavioral inventory were assumed to be equivalent in difficulty and sensitivity with $b_k = 1.5$ and $s_k = 1$ for all k . The item parameters were selected to generate sum score values with distributions resembling those in some of our own data. Ideally,

**FIGURE 2**

Histograms of simulated latent trait and checklist sum scores for outcome behavior environmental covariate ($N = 20,000$).

the items should be selected to represent different thresholds across the entire practical range of the latent trait (see, e.g., Lord & Novick, 1968). In practice, however, this is seldom possible for relatively infrequent symptoms of abnormal behavior and environmental indicators (see, e.g., Eaves et al., 2005).

The simulated 0/1 item responses for the environmental covariate and behavioral checklist were added to generate a sum score for the environment and symptom count for each subject. Figure 2a–d shows the distributions of the la-

tent trait scores and checklist sum scores for the behavioral and environmental checklists. The reversed J-shaped distribution is typical of many clinical measures derived from symptom counts.

The matrices of twin correlations for the latent traits and checklist scores are given in Table 2. The correlations for the square roots of the sum scores for the symptoms and environmental items are also included because scale transformation is a typical precursor to statistical analysis in which the effects of ignoring heteroscedasticity are expected to

TABLE 2
Summary of Simulated Data for Outcome Phenotype (P) and Environmental Covariate (E) in 10,000 MZ and 10,000 DZ Twin Pairs

		Correlation ($\times 1,000$ MZ upper triangle, DZ lower triangle)												
		Raw				Sum				Root (Sum)				
		Twin 1 (T1)		Twin 2 (T2)		Twin 1 (T1)		Twin 2 (T2)		Twin 1 (T1)		Twin 2 (T2)		
Score		E	P	E	P	E	P	E	P	E	P	E	P	
Raw	T1	E		495	748	375	799	435	598	336	823	454	616	343
		P	509		372	687	393	888	289	608	406	922	304	627
	T2	E	744	380	499	590	329	801	441	608	343	826	454	
		P	373	444	501	295	614	395	890	305	632	410	921	
Sum	T1	E	804	406	601	299	373	558	281	942	369	534	278	
		P	455	889	334	386	394	271	598	372	946	278	578	
	T2	E	594	293	803	393	564	275	379	535	274	941	368	
		P	326	393	440	890	269	363	375	282	577	379	947	
Root	T1	E	826	413	616	307	941	386	536	268	380	535	285	
		P	473	922	348	403	387	946	274	362	392	287	586	
	T2	E	608	303	826	410	537	274	942	378	536	280	377	
		P	343	408	462	922	275	358	369	946	281	372	382	
Mean		-0.017	-0.012	-0.017	-0.008	1.760	5.950	1.753	5.921	0.945	1.999	0.940	1.994	
Median		-0.009	-0.011	-0.021	-0.010	1.000	4.000	1.000	4.000	1.000	2.000	1.000	2.000	
SD		0.994	1.004	0.993	0.993	2.267	6.149	2.271	6.136	0.931	1.399	0.932	1.395	

Note: Twin correlations for outcome phenotype and environmental covariate shown in bold.

generate apparent interactions. The choice of square-root transformation (see Bartlett, 1947) for this application is suggested by the scoring of items showing relatively infrequent endorsement (e.g., Eaves & Eysenck, 1977). Other scales suggest other transformations (see, e.g., Eaves et al., 1989) but, ultimately, the best approach is to integrate the genetic model for the latent trait and covariate(s) with the psychometric model for how these are projected onto the items chosen to measure them (see below). The twin correlations for the derived scores are lower than those for the latent traits because of the inherent stochastic error attached to each item response.

Analysis of Simulated Data to Detect $G \times E$ Interaction

Regression of Intra-Pair Outcome Variance on Pair Mean Outcome for MZ and DZ Twins

A seminal paper by Jinks and Fulker (1970) pointed out that the expected means of separated MZ twin pairs were estimates of genetic effects and the absolute intra-pair differences were estimates of the effects of random environmental effects. If sensitivity to the environment, measured by intra-pair variability was a function of average genetic deviation of the pair, Jinks and Fulker predicted a significant regression of absolute intra-pair difference (or intra-pair variance) on pair means for separated twins. In principle, the regression of variance on mean need not be linear or monotonic. The same basic approach, used with other types of relative pairs (e.g., MZ and DZ twins reared together), would reveal other kinds of interaction involving the modulation of intra-pair differences by sources of between pair variation (e.g., Eaves & Eysenck, 1976). Figure 3a–c shows the (linear) regression of intra-pair variances on pair means

for the behavioral outcome in MZ and DZ twins: latent trait (Figure 3a), sum scores (Figure 3b), and square-root transformed sum scores (Figure 3c).

As expected, Figure 3a shows little evidence that intra-pair variance changes with pair mean for the simulated latent trait scores, although the regression line for DZ twins is significantly more elevated than that for MZs due the segregation of genetic differences within pairs of DZ twins. Indeed, twice the difference between the elevations of the line for DZs and MZs is an estimate of the genetic variance in the measured trait. By contrast, the sum scores show a markedly different picture with a very steep regression of intra-pair variance on mean for MZ pairs, indicating that the apparent effects of the individual twins' environments are modulated by the sources of differences between pairs, including $G \times E$ interaction if the differences between pairs are partly genetic. The slope for DZs is much steeper than that for MZs, reflecting the apparent increase in the importance of genetic factors as a function of differences between families. Without a deeper examination, it might be tempting to infer that $G \times E$ interaction is responsible for the increase in genetic variance as a function of increasing 'dose' of the environmental covariate. The difference between the apparent absence of interaction in Figure 2a and the very marked interaction in Figure 2b says nothing about biological process but depends entirely on the characteristics of the specific behavioral items chosen to measure it. Other items or other ways of scaling could produce entirely different patterns of interaction. Thus, patterns of interaction may lead no deeper than the theory of measurement, if indeed there is one. The point is emphasized by the fact that a square-root transformation (e.g., Bartlett, 1947) of the sum scores derived from basic statistical considerations removes much of the apparent support for non-additivity (Figure 3c).

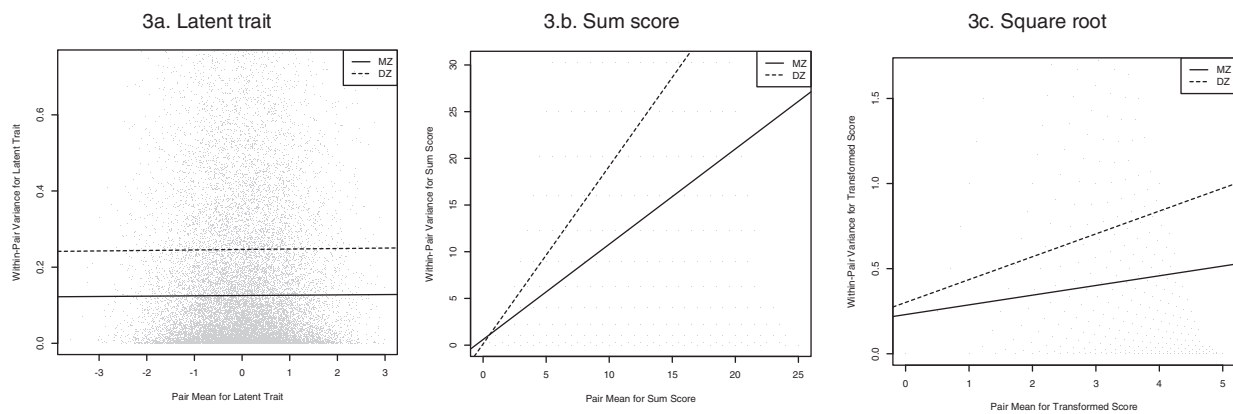


FIGURE 3

Regression of with twin pair variances for latent trait on pair means for trait in MZ and DZ pairs. Note: The top 5% of intrapair differences are omitted from the diagrams (but not from the regressions) to improve scaling of the ordinate.

Regression of Intra-Pair Outcome Variance on Pair Mean Environmental Covariate for MZ and DZ Twins

The above analysis does not specify any specific environmental covariate. The simulated data include a ‘measured’ environmental covariate that is highly correlated between twin pairs and has a substantial linear relationship with the underlying behavioral trait. Figure 4a–c shows the regression of intra-pair outcome variances on pair mean environment for MZ and DZ pairs when behavior and outcome are both assessed by latent trait (Figure 4a), sum score (Figure 4b), and square root transformed sum score (Figure 4c). The graphs now illustrate the modulation of genetic and environmental within-pair variance of MZ and DZ twins by the shared environmental covariate.

The regressions on pair mean environmental covariate resemble those in Figure 3 for the regression on mean phenotype score. There is little or no evidence of any slope in Figure 4a; the lines are flat, with that for DZs being more elevated than that for MZs as a function of the genetic contribution to the outcome. Analysis of sum scores from the behavioral and environmental checklists yields striking support for the increase in genetic contribution as a function of the environmental covariate ($G \times E$, Figure 4b). The same pattern is to be expected with any covariate that has a linear relationship to the latent trait when checklist scores are employed to summarize the behavior of interest. As before, the square-root transformation removes much of the evidence for environmental modulation of genetic and environmental components of variance with twin pairs (Figure 4c).

Fitting Structural Models Allowing for the Environmental Modulation of Genetic and Environmental Path Coefficients

The above analyses are instructive and require nothing more than twin data and the most basic resources for data

summary and graphical display. Such basic considerations are often overlooked, given the availability of efficient software for structural modeling of family resemblance such as Mx (Neale et al., 2006) and openMx (Boker et al., 2011). However, the availability of ready-built code for a variety of models allows the investigator to demonstrate relative sophistication without ever looking critically at how the assumptions of a specific model map onto those used to generate the data to which it is applied.

The dependence between scale of measurement and the results of genetic modeling is revealed by fitting a variety of models to the simulated outcome measures of behavior, scored as a latent trait, checklist sum score, and square-root transformed sum score allowing for the fixed main effect, β , of the measured environmental covariate on outcome, the random effects of genes (a), residual shared environment (c), and individual unique environment (e), and the modulation of a, c, and e by the fixed environmental covariate (parameters γ , η , and δ , respectively). This approach is similar to that proposed by Purcell (2002), treating the covariate as a fixed effect in a manner comparable to those used in the analysis of $G \times E$ in other species (see, e.g., Bucio-Alanis & Hill, 1966; Perkins & Jinks, 1973; and references).

Four models were fitted to each of the three assessments of outcome by full information maximum likelihood (FIML) in openMx (Boker et al., 2011). (1) Estimating the mean, μ , and paths from genes (a), shared environment (c), and unique environment (e) to outcome without any regression of outcome on measured environment (β) or modulation of a, c, and e by the measured environment (i.e., no interaction). (2) Estimating all the parameters of model 1 plus the main effect of the environmental covariate on outcome (β). (3) Estimating the parameters of model 2 plus the modulation (γ) of the genetic path a by the environment ($G \times E$ interaction). (4) A ‘full’ model including the

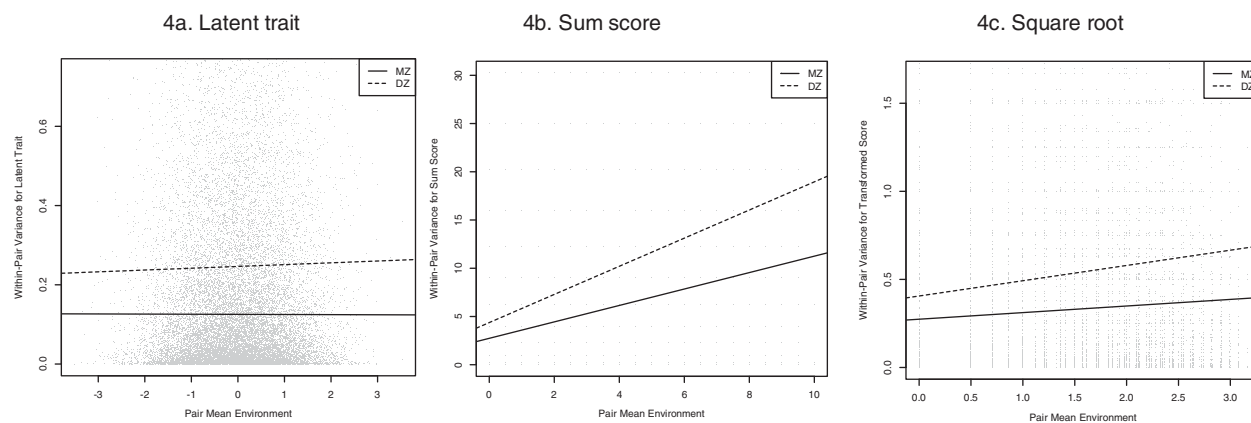


FIGURE 4

Regression of within twin pair variances for outcome on pair means for environmental risk trait in MZ and DZ pairs. Note: The top 5% of inpair differences are omitted from the diagrams (but not from the regressions) to improve scaling of the ordinate.

parameters of model 3 plus the addition modulation of c and e by the measured environment (i.e., adding interaction between the measured environment and the random effects of the shared and unique environment).

Results of fitting the four models to the three pairs of simulated outcomes and environments are summarized in Table 3.

All three measurement scales show strong support for the main effects of genes, environment and covariate on outcome. The estimates of β are all highly significant and the improvement in fit over the model that excludes any effect of covariate is substantial. These are very large samples for a twin study, but even a study one-tenth of the size (1,000 of both MZ and DZ pairs) is expected to yield highly significant main effects of the shared environment. As with the simple variance-mean regressions (see Figure 4), support for environmental modulation of random genetic and environmental effects is contingent on the units chosen to measure behavior. When the latent outcome and environment are both measured directly there is absolutely no gain in support from adding $G \times E$ ($\chi^2(1) = 0.3$) or modulation of the shared and unique environment ($\chi^2(2) = 2.4$). The picture is sharply different when analysis is based on checklist sum scores. There is now highly significant $G \times E$ ($\chi^2(1) = 977.6$) and strong additional support for modulation of c and e ($\chi^2(2) = 362.6$). The evidence for interaction is likely to remain even in samples one-tenth as large or smaller with these effect sizes. Square-root transformation removes all significant support for modulation of genetic and environmental effects by the measured environment. Thus, all evidence for interaction between genes, environment, and an environmental covariate is generated by summarizing behavior by adding responses to checklists of items.

Integrating Models for the Causes of Variation With Theories of Measurement

The above analyses confirm the intimate connection between the model for the effects of genes and environment and the instrument used to assess behavior and its covariates. The conclusions of a genetic analysis are acutely sensitive to the units of measurement when it comes to interpreting $G \times E$ interaction and other kinds of non-additivity at anything other than the most superficial level in the absence of any theory about how individual differences in liability translate into the measures used for analysis.

Such results, long familiar to statistical geneticists, require that claims to detect $G \times E$ for human behavior need to be viewed far more critically than is typical in the current climate. At the very least it is incumbent on those working in the area take the time to evaluate the scientific significance of their findings more carefully by taking into account the theory of measurement that is typically ignored in psychiatric genetics. As a final step, the first 2,500 MZ and DZ pairs (5,000 pairs, in total) of the 200,000 simulated pairs above were employed to demonstrate the kind of approach that investigators could employ to deepen theoretical and substantive understanding of the processes they profess to study. The reduction in sample size significantly reduces computation time without undermining the basic conclusion.

The model comprises two elements: 'genetic' and 'psychometric'. The genetic model comprises the parameters that characterize the additive and non-additive effects of genes and environment on variation in the latent trait (θ) such as that employed above in Table 3. The psychometric model comprises the function that maps onto θ the scores or responses used for assessment. In our example, where the simulated checklist items are all assumed to have the same

TABLE 3
Effects of Scale of Measurement on Tests for Modulation of Genetic Effects by Measured Environment (G × E)

Score		Model			
		No covariate	Main E effect	E + G × E	Full interaction
'True' df	μ	-0.0095 (0.0062)	-0.0014 (0.0053)	-0.0014 (0.0053)	-0.0014 (0.0053)
	a	0.6961 (0.0124)	0.6923(0.0106)	0.6922(0.0107)	0.6896(0.0101)
	c	0.4475 (0.0177)	0.1240 (0.0523)	0.1245 (0.0529)	0.1346 (0.0452)
	e	0.5590 (0.0039)	0.5020 (0.0035)	0.5020 (0.0037)	0.5018 (0.0035)
	β	-	0.5018 (0.0047)	0.5018 (0.0047)	0.5018 (0.0047)
	γ	-	-	0.0021 (0.0041)	0.0103(0.0074)
	η	-	-	-	-0.0368 (0.0266)
	δ	-	-	-	-0.0022 (0.0036)
	-2ln(l)	104,838.3	94,812.5	94,812.2	94,810.6
	χ ²	-	10,025.8	0.3	2.4
	Sum	μ	5.9378 (0.0372)	4.3023(0.0411)	4.2482 (0.0372)
a		4.2930 (0.0573)	4.1645 (0.1390)	3.2990 (0.0500)	3.6805 (0.0913)
c		2.1580 (0.1305)	0.6971 (0.7295)	1.0795 (0.0905)	0.8494 (0.1040)
e		3.8896 (0.0266)	3.8073 (0.0297)	3.7703 (0.0244)	3.3260 (0.0296)
β		-	0.9308 (0.0130)	0.9391 (0.0151)	0.9583 (0.0152)
γ		-	-	0.4153 (0.0147)	0.2224 (0.0213)
η		-	-	-	0.0227 (0.0531)
δ		-	-	-	0.2682 (0.0152)
-2ln(l)		252,904.3	248,084.1	247,106.5	246,743.9
χ ²		-	4,820.2	977.6	362.6
Sqrt (Sum)		μ	1.9969 (0.0085)	1.5176 (0.0102)	1.5179 (0.0101)
	a	0.9100 (0.0215)	0.8946 (0.0215)	0.8782 (0.0219)	0.8597 (0.0254)
	c	0.5592 (0.0305)	0.2792 (0.0590)	0.2803 (0.0553)	0.3064 (0.0606)
	e	0.8994 (0.0062)	0.8882 (0.0062)	0.8881 (0.0061)	0.9003 (0.0091)
	β	-	0.5083 (0.0072)	0.5069 (0.0072)	0.5071 (0.0072)
	γ	-	-	0.0172 (0.0073)	0.0365 (0.0161)
	η	-	-	-	-0.0276 (0.0421)
	δ	-	-	-	-0.0137 (0.0073)
	-2ln(l)	134,536.8	129,854.7	129,849.2	129,845.7
	χ ²	-	4,682.1	5.5	3.5

Note: See Table 4 for definition of symbols. Path coefficients are not standardized to facilitate comparison of estimated effects as a function of adding interactions to model.

endorsement probability, $p(\theta)$, the log-likelihood that a subject with trait value θ endorses r out of N items is given by the binomial theorem:

$$L(r, N, \theta) = r \ln[p(\theta)] + (N - r) \ln[(1 - p(\theta))] + C,$$

and the unconditional likelihood of a specific outcome score, r , is the integral

$$L(r, N) = \int_{-\infty}^{\infty} L(r, N, \theta) d\theta.$$

The analytical task devolves upon estimating the parameters relating $p(\theta)$ to θ , that is, the parameters of the psychometric model and those of the 'genetic' model accounting for the variation in the latent trait θ : the additive genetic and environmental paths, a , c , and e , including the regression of liability on the covariate and the coefficients modulating the main effects of genes and environment, γ , η , and δ . The assumption of equivalent items requires only the estimation of one free parameter of the psychometric model: the item difficulty (threshold), b . The sensitivity, s , is fixed at 1, allowing the paths a , c and e to set the scale of variance in the latent trait.

The model for the trait in the j th twin of the i th pair is thus:

$$\theta_{i,j} = \beta X_{i,j} + (a + \gamma X_{i,j}) A_{i,j} + (c + \eta X_{i,j}) C_{i,j} + (e + \delta X_{i,j}) E_{i,j},$$

where $X_{i,j}$ is the corresponding covariate value. In this application, the latent trait covariate score, not the environmental checklist score, was used to simplify the model. $A_{i,j}$, $C_{i,j}$, and $E_{i,j}$ are the random additive genetic, shared environmental and unique environmental effects standardized to zero mean and unit variance. The correlation between the genetic effects is assumed to be 1 for MZ pairs and $1/2$ for DZs. The correlation between the shared environmental effects of twins is 1 and that for the unique environments is 0.

Estimates of the model parameters may be obtained in a variety of ways. We employed a Markov Chain Monte Carlo approach (MCMC) using the Gibbs sampler (see, e.g., Gilks et al., 1996) implemented in BUGS (Bayesian analysis using Gibbs sampling; Lunn et al., 2012). Elsewhere we have used this approach to estimate parameters of models combining elements of genetic and item-response theory (Eaves et al., 2005) in twin data and for models involving $G \times E$

TABLE 4**MCC Estimates of Parameter (θ) in Equivalent Items IRT Model for Main Effects and Modulation of Genes and Environment on Sum Scores**

Parameter		Statistic					
θ	Description	Mean	SD	Accuracy	2.5% CI	Median	97.5% CI
a	Item difficulty ('threshold')	1.067	0.012	0.0005	1.043	1.067	1.091
b	Item discriminating power (fixed scale factor)	1.000	–	–	–	–	–
a	Main genetic path	0.732	0.018	0.0014	0.691	0.733	0.764
c	Unmeasured shared environmental path	0.018	0.122	0.0133	–0.221	0.026	0.236
e	Residual (non-shared) environmental path	0.533	0.012	0.0004	0.510	0.533	0.556
β	Regression of latent trait on measured environment	0.521	0.012	0.0003	0.499	0.521	0.544
γ	Modulation of genetic effect ('G \times E')	–0.017	0.016	0.0009	–0.050	–0.018	0.015
η	Modulation of shared environmental effect (C \times E)	–0.005	0.058	0.0056	–0.112	–0.004	0.111
δ	Modulation of unique environmental effect (E \times E)	–0.038	0.012	0.0004	–0.062	–0.038	–0.013
r_{DZ}	MZ twin correlation in median measured environment	0.339	0.012	0.0009	0.319	0.337	0.368

interaction and correlation (Coventry et al., 2010; Eaves et al., 2003; Wray et al., 2008). The MCMC approach is especially convenient for fitting models that require large amounts of numerical integration to compute likelihoods and has the added flexibility of yielding estimates of the standard errors and confidence intervals of parameters at little additional computational cost.

Results of fitting the full combined psychometric and non-additive genetic model are summarized in Table 4. Estimates are obtained from every fifth MCMC sample from 30,000 iterations ($N = 6,000$) after a 20,000 iteration 'burn in'. Computations required approximately six hours on a MacBook Pro laptop computer.

The results are striking. Once the theory of measurement is integrated with the genetic model there is absolutely no evidence of G \times E interaction or of any modulation of the shared and unique environmental parameters by the environmental covariate. The 95% confidence intervals of γ, η , and δ all lie either side of zero, and the median estimates of the modulation parameters are all close to zero. Furthermore, the estimates of the main genetic and environmental parameters and the regression of latent trait on measured environment are all close to those used in simulating the original data. In particular, after the effect of the measured environment is extracted, there is no residual contribution of the shared environment to outcome liability ($-0.221 < c < 0.236$) and the estimated residual correlation of MZ twins (0.660) is close to twice that estimated for DZs (0.339), as expected when only additive genetic effects contribute to residual twin resemblance.

Discussion

The psychometric model assumed in the simulation and MCMC analysis is simpler than might apply in practice. Items may not be equivalent and the detailed implications of different methods of environmental assessment have still to be explored. Nevertheless, our simulation captures the essence of the problem and it is unlikely that elaboration of

the measurement model with different item difficulties and sensitivities will substantially change our conclusions here.

Geneticists have long recognized that phenotypic distributions depend on parameters of the underlying genetic model for individual differences. Thus, the moments of distributions in kinship data are well known to depend on factors as diverse as the number of loci affecting a complex trait (e.g., Carmelli et al., 1979; Karlin et al., 1981), their dominant and epistatic interactions (e.g., Fisher et al., 1932), and G \times E interactions (e.g. Jinks and Fulker, 1970). On the whole, geneticists have been far more cautious about the using the properties of observed phenotypic distributions in kinships to infer subtleties of the genetic architecture of complex traits because a variety of more or less arbitrary factors, having little or nothing to do with genetics, can affect the more subtle features of trait distributions. Paramount among such factors are those arising from the fact that the scales used to measure variation have an ill-defined relationship to underlying biological differences such that changes in the units or method of measurement can lead to drastically different conclusions about the genetic architecture of the underlying biological system. Mather and Jinks (1982) offer a classical statement of the interdependence of measurement and genetic inference:

The scale on which the measurements are expressed for the purposes of genetical analysis must therefore be reached by empirical means. Obviously it should be one which facilitates both the analysis of the data and the interpretation and use of the resulting statistics. *The scale should preferably be one on which ... the interactions among the genes and between genotype and environment are absent, or at any rate as small as they can reasonably be made.* (p. 64, italics added)

Lack of attention to this goal leaves in question the heuristic value of claims to find G \times E in psychiatric data.

Mather and Jinks (1982) recommend that 'so far as possible the non-allelic genes and non-heritable agents should all be additive in action' but also caution that such scales may be hard to find since 'each gene and each non-heritable agent may be acting on its own scale' and the elegance of a

parsimonious additive model may be elusive. The problem is that psychiatric geneticists seldom bother to look. We are not blessed with decisions as simple as whether to measure body weight in kilograms or log-kilograms, though even here the choice of scale will not be neutral with respect to conclusions about the contributions of additive and non-additive effects.

With respect to behavior, the issue often comes down to decisions about which constellations of items, combined in which way, best characterize the salient latent behavioral outcomes and psychosocial risk factors of interest. The relationship between the numbers generated by a test and the way genes and environment work is tenuous and theory-dependent. There is an intimate connection between the choice of measure, and conclusions drawn about the relative importance of genes, environment and the various possible interactions between them. Elegant pictures of the role of G × E interaction, such as those in Figures 3 and 4, may be no more robust than the items selected to measure the hypothesized latent variable, the rule used to combine them, or how the scores are scaled after they have been combined. It has long been recognized that as simple a step as taking the square root or arcsine of a personality test score can remove otherwise striking evidence of G × E and sex differences in genetic effects in twin data (e.g., Eaves & Eysenck, 1977; Eaves et al., 1989) and yield a far more parsimonious and heuristically powerful, additive genetic, and environmental model. Indeed, it has been shown that analysis of psychological test scores that ignore the relationship between test and the underlying trait can even simulate the segregation of a single gene of large effect in kinship data (Eaves, 1983).

The recent emphasis on the importance of G × E in psychiatric genetics necessitates a cautionary re-examination of the relationship between measurement and G × E for the benefit of researchers, editors, and others who make critical decisions about the seriousness with which findings about G × E are to be received (see also Duncan & Keller, 2011).

The current academic and funding environment creates pressure to publish novel and more subtle findings. Yet, there may be nowhere where the risks of perpetuating and funding replicable artifact are greater than in attempts to demonstrate G × E for psychiatric phenotypes and risk factors assessed by checklists of items. The above analyses counsel the exercise of greater thoroughness before publishing claims to detect G × E interaction for psychiatric outcomes. Otherwise, statistically significant findings may reveal less about the biology of behavior than they do about cultural incentives to premature publication.

Acknowledgments

This research was supported by grant DA024413 from the U.S. Department of Health and Human Services (PI E Jane Costello, Duke University). I am grateful to Mike Neale,

Tim York, Brad Verhulst, Judy Silberg, Nick Martin, and John Hewitt for the helpful advice and discussion.

References

- Bartlett, M. S. (1947). The use of transformations. *Biometrics*, 3, 39–52.
- Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., ... Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, 76, 306–317.
- Bucio-Alanis, L., & Hill, J. (1966). Environmental and genotype-environmental components of variability. II. Heterozygotes. *Heredity*, 21, 115–127.
- Carmelli, D., Karlin, S., & Williams, R. R. (1979). A class of indices to assess major-gene versus polygenic inheritance of distributive variables. *Progress in Clinical and Biological Research*, 32, 259–270.
- Coventry, W. L., James, M. R., Eaves, L. J., Gordon, S. D., Gillespie, N. A., Ryan, L., ... Wray, N. R. (2010). Do 5HTTLPR and stress interact in risk for depression and suicidality? Item response analyses of a large sample. *American Journal of Medical Genetics, B, Neuropsychiatric Genetics*, 153, 757–765.
- Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, 168, 1041–1049.
- Eaves, L. J. (1983). Errors of inference in the detection of major gene effects on psychological test scores. *Advances in Behavioral Research and Therapy*, 35, 1170–1189.
- Eaves, L., Erkanli, A., Silberg, J., Angold, A., Maes, H. H., & Foley, D. (2005). Application of Bayesian inference using Gibbs sampling to item-response theory modeling of multi-symptom genetic data. *Behavior Genetics*, 35, 765–780.
- Eaves, L. J., & Eysenck, H. J. (1976). Genotype × age interaction of neuroticism. *Behavior Genetics*, 6, 359–362.
- Eaves, L. J., & Eysenck, H. J. (1977). A genotype-environmental model for psychoticism. *Advances in Behavioral Research and Therapy*, 1, 5–26.
- Eaves, L. J., Eysenck, H. J., & Martin, N. G. (1989). *Genes, culture and personality: An empirical approach*. New York, NY: Academic Press.
- Eaves, L. J., Silberg, J., & Erkanli, A. (2003). Resolving multiple epigenetic pathways to adolescent depression. *Journal of Child Psychology and Psychiatry*, 44, 1006–1014.
- Fisher, R. A., Immer, F. R., & Tedin, O. (1932). The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics*, 17, 107–124.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman & Hall.
- 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.

- Karlin, S., Williams, P. T., & Carmelli, D. (1981). Structured exploratory data analysis (SEDA) for determining mode of inheritance of quantitative traits. I. Simulation studies on the effect of background distributions. *American Journal of Human Genetics*, 33, 262–281.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. New York, NY: Addison-Wesley.
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. London: Chapman & Hall.
- Mather, K., & Jinks, J. L. (1982). *Biometrical genetics, the study of continuous variation* (3rd ed). London & New York: Chapman and Hall.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical modeling* (7th ed.). Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.
- Perkins, J. M., & Jinks, J. L. (1973) The assessment and specificity of environmental and genotype-environmental components of variability. *Heredity*, 30, 111–126.
- Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research*, 5, 554–571.
- Wray, N. R., Coventry, W. L., James, M. R., Montgomery, G. W., Eaves, L. J., & Martin, N. G. (2008). Use of monozygotic twins to investigate the relationship between 5HT-TLPR genotype, depression and stressful life events: An application of item response theory. *Novartis Foundation Symposium*, 293, 48–59.
-