Journal of Child Psychology and Psychiatry 44:7 (2003), pp 1006–1014

Resolving multiple epigenetic pathways to adolescent depression

Lindon Eaves,[1] Judy Silberg,[1] and Alaattin Erkanli[2]

[1]Virginia Institute for Psychiatric and Behavioral Genetics, USA; [2]Epidemiology Program, Duke University Medical Center, USA

Background: Genotype × environment interaction (G × E) arises when genes influence sensitivity to the environment. G × E is easily recognized in experimental organisms that permit randomization of genotypes over fixed environmental treatments. Genotype–environment correlation (rGE) arises when genetic effects create or evoke exposure to environmental differences. Simultaneous analysis of G × E and such 'active' or 'evocative' rGE in humans is intractable with linear structural models widely used in behavioral genetics because environments are random effects often correlated with genotype. The causes of the environmental variation, therefore, need to be modeled at the same time as the primary outcome. **Methods:** A Markov Chain Monte Carlo approach is used to resolve three distinct pathways involving genes and life events affecting the development of post-pubertal depression in female twins and its relationship to pre-pubertal anxiety: 1) the main of genes and environment; 2) the interaction of genes and environment (G × E); and 3) genotype–environment correlation (rGE). **Results:** A model including G × E and rGE in addition to the main effects of genes and environment yields significant estimates of the parameters reflecting G × E and rGE. Omission of either G × E or rGE leads to overestimation of the effects of the measured environment and the unique random environment within families. **Conclusions:** 1) Genetic differences in anxiety create later genetic differences in depression; 2) genes that affect early anxiety increase sensitivity (G × E) to adverse life events; 3) genes that increase risk to early anxiety increase exposure to depressogenic environmental influences (rGE). Additional genetic effects, specific to depression, further increase sensitivity to adversity. Failure to take into account the effects of G × E and rGE will lead to misunderstanding how genes and environment affect complex behavior. **Keywords:** Anxiety, depression, life events, genotype × environment interaction, genotype, environment correlation, adolescence, development, Bayesian, Markov Chain Monte Carlo, twins. **Abbreviations:** G × E: Genotype × environment interaction; rGE: Genotype–environment correlation; MZ: monozygotic; DZ: dizygotic; MCMC: Markov Chain Monte Carlo.

The paths from DNA to psychopathology are long and tortuous. Almost certainly, they involve the action, interaction and correlation of many genes and environmental factors whose effects change and/or accumulate through development as a result of endogenous mechanisms and the interplay between the person and the environment. In this paper we try to bring together in a single model three separate strands of previous genetic analysis of adolescent depression that, until now, have only been considered in isolation: interaction between genes and life events in the etiology of depression; genetic influences on life events; and heterotypic (genetic) continuity/comorbidity between early anxiety and later depression. To accomplish this goal, we combine longitudinal, genetically informative data from the Virginia Twin Study for Adolescent Behavioral Development (VTSABD) within a Bayesian framework for statistical analysis that supersedes many earlier methods for the facility with which it can handle problems in non-linear genetic modeling.

There are several different ways in which genes may affect a psychiatric disorder (Kendler & Eaves, 1986; Rutter & Silberg, 2001). Some genes may affect overall liability to disorder ('main effect' of genes). Other genes may affect liability by influencing sensitivity of the individual to environmental factors ('genotype × environment interaction', G × E) creating genetic variation in the regression of phenotype on environment. Still other genes may affect the probability of exposure to environmental risk factors ('genotype–environment correlation', rGE). The classical example of rGE in experimental genetics arises when the offspring phenotype is affected by the maternal genotype. In humans, especially, we recognize also that individuals may create, select or elicit environments that are correlated with their genotype ('active' or 'evocative' rGE; Plomin, Lichtenstein, Pedersen, McClearn, & Nesselroade, 1990; Kendler, Neale, Kessler, Heath, & Eaves, 1993).

Different genes may affect the phenotype through two or three of these pathways simultaneously. Thus, genes that affect overall liability may also increase sensitivity to the environment and influence exposure to the environment (Mather & Jinks, 1982; Jinks & Fulker, 1970). In addition, the expression of genetic and environmental effects may change during development (Eaves, Long, & Heath, 1986) leading to different genes affecting the same phenotype at different ages and/or the same genes having different phenotypic expression at different ages ('heterotypic continuity').

© Association for Child Psychology and Psychiatry, 2003.
Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

Genotype × environment interaction (G × E) is a widespread property of genetic systems (Mather & Jinks, 1982) that arises when sensitivity of the phenotype to environmental influences is partly under genetic control. Sensitivity to the environment may be mediated by different genes from those contributing to the main effects of genetic differences (Caligari & Mather, 1975) and different genes may control sensitivity to different measured aspects of the environment (Mather, 1975).

The significance of G × E and rGE for human behavioral differences and disorders has been widely acknowledged (Cattell, 1965; Jinks & Fulker, 1970; Eaves, Last, Martin, & Jinks, 1977; Scarr & McCartney, 1983), but clear examples of G × E in humans are few (Kendler & Eaves, 1986).

In experimental organisms, it is possible to randomize sets of genotypes over the range of relevant environments so any differences between genetic architecture (e.g., amount of genetic variance) in different environments indicate the presence of G × E interaction. In humans, it is common practice to detect G × E in an analogous way by stratifying genetically informative individuals (e.g., twin pairs) by hypothesized environmental factors and to compare the contributions of genetic effects between strata. However, this approach assumes that the genes affecting the measured trait do not also affect exposure to the salient environment. If there is rGE, stratification by an environmental variable will result in differences between allele frequencies in the different strata and thus produce spurious indication of G × E interaction.

Recognizing this problem has resulted in investigators restricting their analyses of G × E to environments that can be shown to be independent of genotype (i.e., for which there is no detectable rGE). This is the approach used by Heath, Eaves, and Martin (1998) in the detection of interaction between genetic risk to depression and marital status, Silberg et al. (2001) to examine the interaction of life events and genetic risk to adolescent depression, and Caspi et al. (2002) in testing for interaction between allelic differences at the MAO locus and early abuse in risk for conduct disorder. Thus, the analysis of G × E has typically required the absence of rGE. At the same time, many studies claim to have demonstrated genetic effects on exposure to specific environments, notably life events (e.g., Kendler et al., 1993; Silberg et al., 1999). Typically, these studies ignore the effects of G × E interaction.

Correlation between genetic and environmental effects presents a hitherto unresolved complication to the analysis of G × E in humans that is not experienced in the study of experimental organisms. For example, stratification by fixed values of an environmental covariate (Neale & Cardon, 1992) does not deal with the interaction and correlations between random variables needed to characterize G × E in humans. The fact that the environment is often a random variable requires that the analysis of G × E models both the genetic and environmental basis of environment and the outcome at the same time. This has typically not been the case.

In reality, we anticipate that all three mechanisms – main effects of genes and environment, G × E and rGE – will contribute to the development of complex behavioral outcomes. Failure to provide an approach to the genetic analysis of behavior that integrates the three sources of genetic difference in the same model is a barrier to a comprehensive model for the roles of genes and environment in human behavior and risk for psychiatric disorders. The dilemma is simple: current methods detect G × E by assuming there is no rGE, but demonstrate repeatedly that many of the most salient environmental factors are correlated with genetic difference (i.e., that there is considerable rGE).

The empirical groundwork for our attempt to integrate these three mechanisms for the action of genes in the development of depression is laid in three earlier papers by Silberg et al. (1999, 2001, 2001a). Silberg, Rutter, and Eaves (2001) showed how the *main effect* of genetic differences in anxiety before age 14 accounted for most of the genetic differences in post-pubertal depression, thus revealing an underlying genetic basis to heterotypic continuity of anxiety and depression. Silberg, Neale, Rutter, and Eaves (2001) showed how the genetic variance in post-pubertal depression increased as a function of increasing environmental stress (*genotype × environment interaction*), using life events selected to be *independent* of genetic differences, notwithstanding the fact that the genes influencing depression are also implicated in exposure to *dependent* life events (Silberg et al., 1999, *genotype–environment correlation*). However, in spite of repeated demonstrations that genetic differences may explain part of the variance in life events, no method has been devised to integrate G × E and rGE in a single analysis. The fact that these papers initially treated the three mechanisms separately is a reflection of the lack of a general framework for their integration.

Our report remedies this deficiency with an approach that we anticipate will have wider application. We show how a Markov Chain Monte Carlo (MCMC) approach (Gilks, Richardson, & Spiegelhalter, 1996; Hastings, 1970; Zeger & Karim, 1991) can resolve the random main effects and interactions of genes and correlated measured life events in the etiology of depression in post-pubertal twin girls.

Materials and methods

Sample and measures

The data comprise assessments of pre-pubertal anxiety and post-pubertal depression and life events in a

sample of adolescent female twin pairs ($N = 467$ MZ, 220 DZ pairs) from the longitudinal Virginia Twin Study of Adolescent Behavioral Development (Meyer, Silberg, Simonoff, Kendler, & Hewitt, 1996; Hewitt et al., 1997; Eaves et al., 1997; Simonoff et al., 1997). DSM-IIIR symptoms of overanxious disorder and depressive disorder were assessed by face-to-face semi-structured psychiatric interview with each child using the Child and Adolescent Psychiatric Assessment (C-CAPA; Angold et al., 1995). Anxiety and depression were both summarized by total count of DSM-IIIR symptoms attributed in the three months prior to interview. Life events in the year prior to interview were assessed by self-report questionnaire completed by twins' mothers (Johnson, 1986) about each of the twins at the time of home visit. Life events associated with post-pubertal depression were selected by preliminary statistical screening using a data mining approach ('MARS'; Friedman, 1991) that employs cross-validation techniques to minimize the chance of false positive conclusions when reviewing large numbers of covariates. Events were selected only for their association with depression and not for independence from genetic influences. The events selected were aggregated into the 'life events score' used in this study: parent becoming less interested or less loving with her; serious illness or injury to herself; breaking up with someone she had been dating regularly; miscarriage or abortion; making failing grades on a report card; parents divorced or separated; death of a close friend; entry into the home of a new partner for mother or father; parent getting into trouble with the law; brother or sister (or stepbrother/stepsister) leaving home.

Model for G × E and rGE

We consider the joint distribution of three random variables in pairs of twins: post-pubertal depressive symptoms, D; adverse life events, E, and pre-pubertal anxiety, A. We let D_{ij}, A_{ij} and E_{ij} be the measures for the jth twin of the ith pair respectively. Our model for depression allows for the main effects of life events, the main effects of genes and the interaction of genes and life events that may correlate with genotype. The model is summarized graphically in Figure 1. Genetic effects are divided into those shared with earlier anxiety and those having effects specific to later depression. Both may interact with life events to modulate the risk for depression.

Some basic features of the model may be inferred from the summary statistics provided in Table 1.

Consider early anxiety first. Symptoms of pre-pubertal anxiety, A, show a small genetic effect since the DZ correlation is significantly less than the MZ correlation (Table 1). In our model, anxiety is an index primarily of early genetic differences that may influence later depression in several ways. For the anxiety score of the jth twin of the ith DZ pair we write:

$$A_{\mathrm{DZ}ij} = \mu_{\mathrm{a}} + \dot{\exists}_{\mathrm{a}}\, g_{\mathrm{a}i,j} + s_{i,j}. \qquad (1)$$

The mean pre-pubertal anxiety is μ_{a}. The within-family environmental deviation of the jth twin of the ith pair is $s_{i,j}$ ($\mathrm{N}[0,\Phi_{\mathrm{a}}^2]$). Genetic effects on pre-pubertal anxiety, $g_{\mathrm{a}i,j}$, may affect depression in three ways: through their main effect on depression and their impact on sensitivity to life events (G × E) as specified in equation 3 below and through their impact on exposure to life

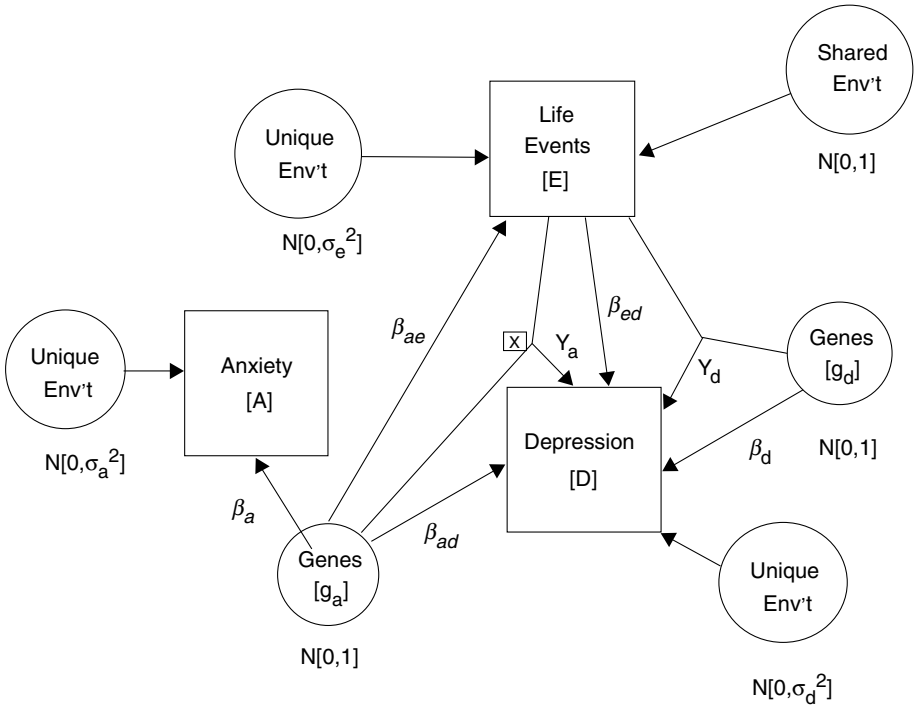

Figure 1 Principal features of model for main effects of genes (G) and life events (E), G × E interaction and genotype–environment correlation on post-pubertal depression in girls. Note: parameter names correspond to those used in text. X denotes G × E interaction pathway. Measured variables (D, E, A) represented by squares, latent variables by circles. Means of measured variables are not included in the figure

Table 1 Correlations for longitudinal data on anxiety, depression and life events in adolescent twin girls

Variables	Correlation	*N* (pairs)
Anxiety–Depression	.177	398
Anxiety–Life events	.109	405
Depression–Life events	.222	702
MZ Depression	.243	243
MZ Anxiety	.243	333
MZ Life Events	.699	263
DZ Depression	.002	95
DZ Anxiety	.126	163
DZ Life Events	.552	99

Note: Anxiety is assessed before puberty. Depression and Life events are assessed after puberty. The MCMC analysis automatically imputes missing values so incomplete response vectors are incorporated in the analysis.

events (G–E correlation) through equation 2. There are numerous ways of parameterizing the model in (1) and subsequent equations. We assume that $g_{ai,j}$ is N[0,1] and the coefficient $\exists_a$ is the path from genotype to anxiety phenotype. The same equation may be used for MZ pairs, recognizing that the twin covariance for genetic effects is greater than that for DZs. In our application we assume that all the main effects of genes are additive: i.e., that the expected covariance of genetic effects in MZs is exactly twice that in DZs (Eaves, 1982).

The model for twin resemblance in life events is also simple. The DZ correlation for the life events scale is large and greater than half the MZ correlation consistent (see Table 1) with substantial shared environmental effects on life events (Eaves, 1982). However, the MZ correlation is still greater, implying that genetic factors influence exposure to life events (genotype–environment correlation). We postulate that the salient environments are influenced by genes that exercise a common effect on anxiety and depression ($g_{ai,j}$ see eq. 1). Thus, the life events score of the *j*th twin of the *i*th pair may be represented by:

$$E_{i,j} = \mu_e + \exists_e c_i + \exists_{ae} g_{ai,j} + e_{i,j,} \qquad (2)$$

where μ_e is the mean life events score, c_i (N[0,1]) is the deviation of the shared family environment of the *i*th family from the population mean and e_{ij} (N[0, Φ_e^2]) is the (environmental) deviation of the *j*th twin of the *i*th pair. The coefficient $\exists_{ae}$ is the path from genes that influence anxiety, $g_{a,}$ to life events. It will be zero if there is no genotype–environment correlation for life events. The path $\exists_e$ reflects the contribution of the shared environment, *c*, to life events.

For depression the model is more complex, allowing for the main effects of genes and life events on depression and G × E interaction. The main effects of genes and their effects in G × E may be divided into two kinds: those genes whose effects are specific to depression and those genes whose effects also influence earlier anxiety. Thus, for DZ twins we write:

$$\begin{aligned} D_{DZij} = \mu_d + \exists_d g_{di,j} + \exists_{ad} g_{ai,j} + \exists_{ed} E_{i,j} \\ + \gamma_d g_{di,j} E_{i,j} + \gamma_a g_{ai,j} E_{i,j} d_{i,j}. \end{aligned} \qquad (3)$$

The mean is μ_d. g_d comprise the genetic effects (N[0,1]) specific to depression and uncorrelated with those affecting *A*. Coefficient $\exists_d$ assesses the direct impact of these genes on depression. Coefficient $\exists_{ad}$ measures the impact on depression of genes with primary effects, g_a, (N[0,1]) on *A*. The main effect of life events on depression is reflected in $\exists_{ed}$. Measurement errors and random environmental differences within families are subsumed in the error terms $d_{i,j}$, (N[0, Φ_d^2]). Multiplicative G × E interaction is represented by the coefficients γ_d and γ_a reflecting respectively the varying sensitivity to the environment arising from genes that primarily affect *D* only (g_d) and differences in sensitivity caused by genes (g_a) primarily affecting *A*. In the absence of G × E both coefficients are zero. A similar expression may be written for MZ twins. The model (3) may be elaborated or simplified in a number of ways. Our model is dictated partly by effects that reflect the pattern of MZ and DZ correlations for the three variables.

Statistical method

Models involving G × E are non-linear but the prevailing software for the statistical genetic analysis of family data relies on the mathematical simplifications of the linear model.

The dominant statistical paradigm, in the analysis of twin data, consists of maximizing the likelihood of the observed data as a function of unobserved, latent, variables that are assumed to reflect the individual genetic and environment effects of family members (e.g., Martin & Eaves, 1977; Neale & Cardon, 1992).

Thus, if we denote the phenotypes of a pair of twins by T_1 and T_2 respectively, we may compute the likelihood of the pair for any latent genetic and environmental effects of the twins, (g_1,g_2,e_1,e_2). We may denote this likelihood by $l[T_1,T_2\,|\,(g_1,g_2,e_1,e_2)]$.

In the typical genetic model, the latent variables, $g_1 \ldots e_2$, are unknown. If, however, the distribution of the latent variables is known, for example if they are multivariate normal, the likelihood is the integral over all possible values of $g_1 \ldots e_2$ of the likelihoods weighted by the probability of occurrence of each set of latent values, i.e., we evaluate the integral of

$$L[T_1, T_2|(g_1, g_2, e_1, e_2)]W(g_1, g_2, e_1, e_2)$$

over all possible values of the four latent variables, g_1,g_2,e_1,e_2.

In the most commonly considered case, in which the phenotypes T_1 and T_2 are linear functions of $g_1 \ldots e_2$ and the distribution of the latent variables is multivariate normal, the integral has an explicit 'one line' algebraic solution (see, e.g., Neale & Cardon, 1992) that can be evaluated in terms of the phenotypic values and the means and covariance structure of the latent variables. This means that the computational demands of maximizing the likelihood with respect to the parameters of a linear model for the phenotype, though still not trivial, are relatively manageable and have been implemented in a number of user-friendly packages for linear structural modeling such as Mx (Neale, Boker, Xie, & Maes, 1999).

In general, there may be no exact algebraic solution to this integral, so it may be necessary to approximate it by choosing a number of sets of values for the latent variables and add them with appropriate weights. A variety of different algorithms have been evolved to

guide the selection of values for the latent variables and appropriate weights. The numerical approximation, however, becomes very computer intensive as the number of latent variables (dimensions) increases since, if p points are required to approximate a given integral in one dimension, we will typically need p^k to approximate the integral in k dimensions. If $p = 10$ and $k = 4$ that means that 10,000 evaluations of the likelihood conditional on values of $g_1 \ldots e_2$ are required to approximate the integral for a single pair of twins. This is not impossible, but is potentially very tedious in the most general case, since many evaluations of the overall likelihood will be required to estimate the parameters of the distribution $W(g_1,g_2,e_1,e_2)$.

In the restricted case in which the environment is measured exactly and independent of genotype, conditioning the expected covariances between relatives on the environmental measures allows for the effects of $G \times E$ to be captured by the linear model, with each family (e.g., twin pair) being sampled from a population with covariance matrix conditional on the environmental values. The test for $G \times E$ amounts to a test for the heterogeneity of genetic components over environmental strata. However, this approach will not work in general, because the measures of the environment may be correlated with genotype or themselves only indices of a latent random variable whose effects are assessed more or less unreliably. Under these circumstances, although the likelihood can easily be written, the integral does not reduce to the same simple form that applies under the linear model.

These facts alone place a premium on methods of model-fitting that do not require tedious coding or large numbers of function evaluations for successful parameter estimation. If there were no other theoretical advantage, the practical benefits alone might justify our considering the alternative Markov Chain Monte Carlo approach (MCMC) to model-fitting. More details of the approach and some applications to non-linear genetic models are given by Eaves and Erkanli (in press).

At one level, we may regard MCMC as an alternative approach to the use of Monte Carlo methods of numerical integration to evaluate likelihoods of observations. However, the resemblance is superficial. MCMC is typically embedded in a Bayesian approach to modeling. ML seeks the parameter values, P, that maximize the likelihood of the data, D, given the parameters: i.e., $L(D|P)$. The Bayesian approach seeks the posterior distribution of the parameters Q, conditional on the data, i.e., $H(Q|D)$. In the Bayesian approach, the individual values of the latent variables *and* the means and components of covariance and/or regression coefficients are all parameters in the model. Starting with an assumed prior distribution of the parameters in Q (which may assume very little apart from the form of the distribution), MCMC simulates a chain of parameter values (including values of individual latent variables) such that, under certain conditions, the distribution of successive sets of simulated values converges to the required distribution of the desired parameters. Statistics derived from many repeated samples for the posterior distribution can be used to yield estimates of the parameter means, variances, quantiles etc.

Many methods have been devised for simulating such a chain of values. However, one very flexible approach is the Gibbs sampler (Gilks et al., 1996). This is the approach to MCMC modeling implemented in the package WinBUGS by the MRC BUGS project in Cambridge, England (Spiegelhalter, Thomas, & Best, 2000). Although implementation of models such as ours require some programming, for the most part it is no more daunting than writing code in Mx or S-Plus, and far less inhibiting that writing FORTRAN code even with the assistance of good software for numerical integration, differentiation and non-linear optimization. The benefits of the approach lie in our ability to specify non-linear models such as ours and to obtain estimates of individual values on latent variables and the sampling distributions of model parameters. Gilks et al. (1996) provide a basic introduction to MCMC methods and many examples of their application. Eaves and Erkanli (in press) provide examples of WinBUGS code for some non-linear models for twin data.

In ML, one function evaluation may be thought of as the computation of the likelihood over a large number values of the latent variables chosen for numerical reasons at a given set of values for the model parameters. In MCMC, an iteration is the computation of the likelihood at a set of latent trait and model parameter values simulated from the current posterior distribution of all the parameters, including the trait values.

Although the method is computer-intensive, it has many advantages over approaches that rely on numerical maximization of the likelihood, especially for non-linear latent variable models. Side-benefits of the MCMC approach are simultaneous estimates of functions of the unknown parameters (means, standard deviations, percentiles, etc.) and automatic imputation of missing and latent variables (Gilks et al., 1996). We assumed that the distribution of all unknown constants is multivariate normal, and that variance components each followed the gamma distribution. WinBUGs was used to generate a chain of 35,000 updates of the Gibbs sampler under the full model allowing for the main effects and interactions of the random genetic and environmental factors underlying *D, A* and *E*. We assumed an uninformative prior distribution of the coefficients and variance components. The first 25,000 samples were discarded for 'burn in' and the next 10,000 used to characterize the distribution of the model parameters conditional upon the data. Several reduced models may also be fitted. We focus on two as critical for the interpretation of our data and previous findings. The first reduced model omitted the parameters relating to $G \times E$ interaction, thus specifying a classical linear model for the main effects of genes and environment on the three measures. The second reduced model also omitted the path from genetic effects on anxiety to life events (genotype–environment correlation). For these models, we sampled 10,000 MCMC updates after a 15,000 cycle burn-in.

Results

Parameter estimates under the full model

Table 2 summarizes the distribution of the means, path coefficients and residual components of

Table 2 MCMC estimates (10,000 updates after 25,000 iteration 'burn in') and summary statistics for model including G × E interaction and rGE on post-pubertal depression in girl twins

Parameter	Description	Mean	SD	2.5%-ile	Median	97.5%-ile
Means						
μ_d	Depression mean	−.0382	.0465	−.1298	−.0375	.0515
μ_a	Pre-pubertal anxiety mean	.0022	.0339	−.0639	.0025	.0678
μ_e	Life events mean	.0008	.0471	−.0917	.0015	.0921
Path coefficients: main effects						
$\exists_d$	Genetic, specific to depression	.4136	.0665	.2817	.4146	.5424
$\dot{\exists}_a$	Genetic, anxiety (may also affect depression)	.4823	.0573	.3723	.4848	.5785
$\dot{\exists}_{ad}$	Genetic, from 'anxiety genes' to depression	.2939	.0854	.1064	.2990	.4475
$\dot{\exists}_{ed}$	Life events to depression	.1278	.0628	−.0005	.1279	.2485
$\dot{\exists}_c$	Shared environment, life events	.7646	.0484	.6658	.7659	.8548
$\dot{\exists}_{ae}$	'Anxiety genes' to life events (G–E correlation)	.3274	.1026	.1161	.3322	.5169
Path Coefficients: G × E interaction						
γ_d	Specific 'depression genes' × life events	.5141	.0679	.3776	.5144	.6483
$\dot{\gamma}_d$	'Anxiety genes' × life events	.2409	.0941	.0581	.2430	.4201
Residual (non-shared) environmental variance components						
Φ_d^2	Unique environment, depression	.4897	.0347	.4261	.4878	.5611
Φ_a^2	Unique environment, anxiety	.7732	.0519	.6782	.7711	.8814
Φ_e^2	Unique environment, life events	.3346	.0264	.2866	.3337	.3891
−2ln(l)	Deviance	5452.0	75.57	5308.0	5450.0	5605.0

variance of the full model including rGE and G × E. The empirical upper and lower 2.5% confidence intervals show that most of the components in the model are statistically significant, many highly so. The coefficients testing for G × E are both significant, showing that genetic effects on anxiety also increase sensitivity to environmental adversity. Of no less importance is the highly significant contribution to G × E from genes that have effects only on depression. Previous analyses of the genetic relationship between anxiety and depression that have not included G × E have not found substantial genetic effects on depression that do not also affect anxiety (Kendler, Heath, Martin, & Eaves, 1986; Kendler, Neale, Kessler, Heath, & Eaves, 1992). The genetic correlation between juvenile anxiety and depression is also large, but there is some evidence of genetic effects specific to depression (Thapar & McGuffin, 1997). One possible explanation of the lower genetic correlation in juveniles could be the non-linearity of the genetic relationship between anxiety and depression resulting from G × E interaction. We find a highly significant path (.3274) from genetic effects on anxiety to exposure to adverse life events (rGE), although the direct main effect of life events on depression (.1278) is relatively small and barely significant. The model thus implies that the major impact of life events on depression arises in those individuals who are especially vulnerable genetically to environmental adversity.

Estimates under reduced models

What happens if we ignore the effects of G × E? Results for two reduced models are summarized in Table 3. The likelihood deteriorates substantially when G × E is removed from the model and is still worse when genotype–environment correlation is also deleted. The fact that MCMC simulates large numbers of individual latent trait values means that there is no exact number of df available for likelihood ratio tests. The small standard errors attached to

Table 3 MCMC estimates (10,000 updates after 55,000 iteration 'burn in') and summary statistics for two reduced models for post-pubertal depression in girl twins

		Omitting G × E		Omitting G × E and rGE	
Parameter	Reduced Model Description	Mean	SD	Mean	SD
$\exists_d$	Genetic, specific to depression	.2589	.1338	.2378	.1252
$\dot{\exists}_a$	Genetic, anxiety (may also affect depression)	.3457	.0365	.3288	.0378
$\dot{\exists}_{ad}$	Genetic, from 'anxiety genes' to depression	.2585	.0753	.2727	.0643
$\dot{\exists}_{ed}$	Life events to depression	.1205	.0578	.2171	.0393
$\dot{\exists}_c$	Shared environment, life events	.7670	.0464	.8107	.0379
$\dot{\exists}_{ae}$	'Anxiety genes' to life events (G–E correlation)	.2358	.0687	–	–
$\dot{\Phi}_d^2$	Unique environment, depression	.7539	.0606	.7417	.0615
Φ_a^2	Unique environment, anxiety	.7696	.0499	.7883	.0615
Φ_e^2	Unique environment, life events	.3212	.0266	.3467	.0257
−2ln(l)	Deviance	5756.0	79.25	5825.0	79.33

Note: Means omitted for simplicity.

these effects in the full model confirms they cannot be ignored.

The reduced models demonstrate the misleading conclusions that might ensue from the typical practice of fitting a purely linear model without G × E when G × E interaction is actually present. Firstly, the effects of genes specific to depression, $\dot{\exists}_{d}$, are less apparent and might even be discounted as not significant. The model including G × E assigns approximately 17% ± 5% of the total variance in depression to genes that affect depression but not anxiety. If G × E is ignored, the proportion falls to 7–8% ± 6%. Secondly, the contribution of unidentifiable non-shared environmental effects, $\Phi^{2}_{d,}$ is grossly overestimated (.7539) in the additive model (Table 3) in comparison with the estimate (.4897) when G × E is included in the model (Table 2). Thirdly, if genetic effects on exposure to life events (rGE) are ignored (setting $\dot{\exists}_{ae}$ to zero), the estimated direct main effect of life events on variance in depression is more than doubled from a barely significant 2.1% ± 1.7% under the full model to 4.9% ± 1.7% when rGE is ignored. The path $\dot{\exists}_{ed}$ is increased from .1278 to .2171.

Discussion and conclusions

The pathway to post-adolescent depression involves the main effects of genes and environment, G × E interaction and G–E correlation. We believe this is the first analysis that has demonstrated the possibility of analyzing G × E in humans when genetic differences also influence exposure to salient random environments (rGE). The analysis was accomplished easily within the MCMC framework. Our results apply to outcomes measured as symptom counts and environments assessed by a count of salient life events. We recognize that the importance of G × E interaction depends on the scale of measurement and may often be removed by a change of scale (Mather & Jinks, 1982). The outcome measures are not normal, although we have assumed that all random effects are normal, so the possibility that some of the G × E is scale-dependent cannot be discounted.

Our integrated model yields a picture of three distinct pathways through which genetic differences affecting early anxiety influence later depression: 1) genes influencing early anxiety affect overall liability of the child to develop depression (genetic 'main effect'); 2) individuals at genetically high risk for anxiety are exposed disproportionately to environmental adversity (G–E correlation); 3) individuals with higher genetic liability, who are exposed to the double disadvantage of correlated environmental adversity, are more sensitive to the damaging effects of their environment (G × E interaction). In addition, when G × E is included in the model, we find that part of the genetic risk to depression is not shared with earlier anxiety, but reflects genes whose effects are specific to depression that have a large effect on sensitivity to environmental stress. These effects would have remained undetected by conventional linear structural models.

We cannot know whether our findings will stand the test of replication, but we hope they will stimulate other investigators in developmental behavioral and psychiatric genetics to experiment with nonlinear models. We also hope that our analysis will help make it clear that 'finding the genes', though important, is not the only positive contribution that genetics can make to analyzing the mechanisms underlying behavioral development. Our analysis of anxiety and depression shows the same genes have different effects at different stages of development and provides a genetic basis for one form of heterotypic continuity and comorbidity. We also show how models for development that ignore the epigenetic interplay between genes and environment in the form of G × E and rGE do not give an adequate account of the development of depression. Indeed, we find that including G × E in the model for depression explains some of what traditional analyses had consigned to the relatively large effect of the 'non-shared' environment. Whatever may be the ultimate outcome of attempts to identify specific genes (quantitative trait loci) that account for the genetic component of depression, a thorough understanding of psychopathology will require that psychiatric geneticists and genetic epidemiologists take seriously the epigenetic mechanisms that depend on the behavior of the human organism and opportunities for learning made possible by human evolution. Our statistical analysis shows that failure to include G × E and rGE in the genetic analysis of depression trivializes the role of the person in his/her own development, leads to overestimation of the contribution of within-family ('non-shared') environmental effects, and overestimates the direct effects of environmental covariates on the depressive phenotype. We hope that our approach may help other investigators to broaden their conceptual and analytical horizons beyond the limitations imposed by current approaches to behavior-genetic modeling.

Acknowledgments

We thank Michael Rutter for helpful discussion. Supported by NIMH grants MH452658 (P.I. L.J.E.), MH55557, MH62368 (P.I. J.L.S.); MH57761 (P.I. Adrian Angold) and by the Carman Trust (P.I. J.L.S.).

Correspondence to

Judy Silberg, Virginia Institute for Psychiatric and Behavioral Genetics, Department of Human Genetics, P.O. Box 98003, Virginia Commonwealth University School of Medicine, Richmond, Virginia, 23298-0003, USA; Email: jsilberg@hsc.vcu.edu

References

Angold, A., Prendergast, M., Cox, A., Harrington, R., Simonoff, E.S., & Rutter, M. (1995). The Child and Adolescent Psychiatric Assessment (CAPA). *Psychological Medicine*, *25*, 739–753.

Caligari, P.D.S., & Mather, K. (1975). Genotype–environment interaction. III. Interactions in *Drosophila melanogaster*. *Proceedings of the Royal Society of London, B 208*, 163–187.

Caspi, A., McClay, J., Moffitt, T.E., Mill, J., Martin, J., Craig, I.W., Taylor, A., & Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*, 851–854.

Cattell, R.B. (1965). Methodological and conceptual advances in evaluating hereditary and environmental influences and their interaction. In S.G. Vandenberg (Ed.), *Methods and goals in human behavior genetics* (pp. 95–140). New York: Academic Press.

Eaves, L.J. (1982). The utility of twins. In V.E. Anderson, W.A. Hauser, J.K. Penry, & C.F. Sing (Eds.), *Genetic basis of the epilepsies* (pp. 249–276). New York: Raven Press.

Eaves, L.J., & Erkanli, A. (2002, in press). Markov Chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and G × E interaction. *Behavior Genetics*.

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

Eaves, L.J., Long, J.C., & Heath, A.C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behavioral Genetics*, *16*, 143–162.

Eaves, L.J., Silberg, J.L., Meyer, J.M., Maes, H.H., Simonoff, E., Pickles, A.R., Rutter, M.L., Neale, M.C., Reynolds, C.A., Erickson, M.T., Heath, A.C., Loeber, R., Truett, K.R., & Hewitt, J.K. (1997). Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *Journal of Child Psychology and Psychiatry*, *38*, 965–980.

Friedman, J.H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, *19*, 1–141.

Gilks, W.R., Richardson, S., & Spielgelhalter, D. (1996). *Markov Chain Monte Carlo in practice*. London: Chapman and Hall.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97–109.

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.

Hewitt, J.K., Silberg, J.L., Rutter, M.L., Simonoff, E., Meyer, J.M., Maes, H.H., Pickles, A.R., Neale, M.C., Loeber, R., Erickson, M.T., Kendler, K.S., Heath, A.C., Truett, K.R., Reynolds, C.A., & Eaves, L.J. (1997). Genetics and developmental psychopathology: 1. Phenotypic assessment in The Virginia Twin Study of Adolescent Behavioral Development. *Journal of Child Psychology and Psychiatry*, *38*, 943–963.

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.

Johnson, J.H. (1986). *Life events as stressors in childhood and adolescence*. Beverly Hills, Sage.

Kendler, K.S., & Eaves, L.J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *American Journal of Psychiatry*, *143*, 279–289.

Kendler, K.S., Heath, A.C., Martin, N.G., & Eaves, L.J. (1986). Symptoms of anxiety and depression in a volunteer twin population: The etiologic role of genetic and environmental factors. *Archives of General Psychiatry*, *43*, 213–221.

Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C., & Eaves, L.J. (1992). Major depression and generalized anxiety disorder. Same genes (partly) different environments? *Archives of General Psychiatry*, *49*, 716–722.

Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C., & Eaves, L.J. (1993). A twin study of recent life events and difficulties. *Archives of General Psychiatry*, *50*, 789–796.

Martin, N.G., & Eaves, L.J. (1977). The genetical analysis of covariance structure. *Heredity*, *38*, 79–95.

Mather, K., & Jinks, J.L. (1982). *Biometrical genetics: The study of continuous variation*. London: Chapman and Hall.

Mather, K. (1975). Genotype × environment interaction II. Some genetical considerations. *Heredity*, *35*, 31–53.

Meyer, J.M., Silberg, J.L., Simonoff, E., Kendler, K.S., & Hewitt, J.K. (1996). The Virginia Twin-Family Study of Adolescent Behavioral Development: Assessing sample biases in demographic correlates of psychopathology. *Psychological Medicine*, *26*, 1119–1133.

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

Neale, M.C., Boker, S.M., Xie, G., & Maes, H.H. (1999). *Mx: Statistical modeling* (5th edn). VCU Box 900126, Richmond, VA 23298: Department of Psychiatry.

Plomin, R., Lichtenstein, P., Pedersen, N., McClearn, G.E., & Nesselroade, J.R. (1990). Genetic influences on life events during the last half of the life span. *Psychology of Aging*, *5*, 25–30.

Rutter, M.L., & Silberg, J.L. (2001). Gene-environment interplay in relative to emotional and behavioral disturbance. *Annual Review of Psychology*, *53*, 463–490.

Scarr, S., & McCartney, K. (1983). How people make their own environments. A theory of genotype–environment effects. *Child Development*, *54*, 424–435.

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

Silberg, J., Pickles, A., Rutter, M., Hewitt, J., Simonoff, E., Maes, H., Carbonneau, R., Murrelle, L., Foley, D., & Eaves, L. (1999). The influence of genetic factors and life stress on depression in adolescent girls. *Archives of General Psychiatry*, *56*, 225–232.

Silberg, J., Rutter, M., & Eaves, L.J. (2001a). Genetic and environmental influences on the temporal

association between early anxiety and later depression in girls. *Biological Psychiatry*, *49*, 1040–1049.

Simonoff, E., Pickles, A.R., Meyer, J.M., Silberg, J.L., Maes, H.H., Loeber, R., Rutter, M.L., Hewitt, J.K., & Eaves, L.J. (1997). The Virginia Twin Study of Adolescent Behavioral Development: Influences of age, gender and impairment on rates of disorder. *Archives of General Psychiatry*, *54*, 801–808.

Spielgelhalter, D.J., Thomas, A., & Best, N. (2000). *WinBUGS version 1.3 User Manual*. MRC Biolstatistics Unit, Cambridge.

Thapar, A., & McGuffin, P. (1997). Anxiety and depressive symptoms in childhood: A genetic study of comorbidity. *Journal of Child Psychology and Psychiatry*, *38*, 651–656.

Zeger, S.L., & Karim, M.R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, *86*, 79–86.

Manuscript accepted 11 March 2003

Copyright © 2003 EBSCO Publishing