Behavior Genetics, Vol. 17, No. 4, 1987

Testing Genetic Models for Multiple Symptoms: An Application to the Genetic Analysis of Liability to Depression

L. J. Eaves,[1] N. G. Martin,[1] A. C. Heath,[1] and K. S. Kendler[1]

Received 15 June 1986—Final 4 Aug. 1986

A model is presented which allows for the contribution of genes and environment to categorical data on multiple symptoms. The model distinguishes between parameters needed to express the relationship between a latent trait and observed responses and the parameters required to represent the causes of variation in the latent trait. The regression of the latent trait on covariates may also be specified. The model is applied to symptoms of depression in 1983 pairs of adult female monozygotic and dizygotic twins. A model which allows only for polygenic variation in the latent trait is supported as well as the "mixed model," which also allows for the effects of a major gene. The likelihood is significantly lower when all genetic effects are ascribed to a single gene. Practical limitations of the method are discussed.

KEY WORDS: latent trait; depression; mixed model; twins; symptoms; major gene; segregation; polygenes; psychometrics; heritability.

INTRODUCTION

Many diagnoses are achieved by combining data on a number of symptoms. This is especially the case with psychiatric disorders, which may be diagnosed on the basis of reponses to a number of specific items in a standardized interview. Similarly, many psychological test scores are as-

This research is supported by Grants AG04954, AA06781, GM32782, GM30250, and MH40828 from the National Institutes of Health. We are indebted to Dr. Greg Carey for his incisive discussion.

[1] Department of Human Genetics and Department of Psychiatry, Medical College of Virginia, Richmond, Virginia 23298.

0001-8244/87/0700-0331$05.00/0

signed to individuals on the basis of their answers to a number of dichotomous items.

Such complex diagnoses and scores may present particular problems for genetic analysis because the categories or scores assigned may not be related in a simple fashion to the primary dimension on which genetic and environmental effects operate. The same basic underlying biological or psychological variable may be assessed more or less arbitrarily by a number of different sets of items. Before embarking on any genetic analysis, therefore, it is necessary to recognize that the items of which our scales are composed can influence the findings of our analysis and that failure to take into account the properties of our measurements may lead to erroneous conclusions about the number and action of genes in the determination of a particular trait. Eaves (1983) has shown, for example, that the effects of a major gene might be inferred incorrectly when no allowance is made for the relationship between the latent dimension on which genetic effects are primarily expressed and the test scores used to summarize behavior. A similar problem is expected to arise when statistical methods such as linear discriminant function analysis are used to generate scores which are then examined directly for evidence of multimodality (e.g., Cloninger *et al.*, 1985).

We present a model for multiitem data which incorporates the distinction between the "genetic" and the "psychometric" aspects of the primary data. Although we describe the model as it would be applied to multiple items of a psychological test, the same approach can be employed in the analysis of any set of multiple-symptom data generated by systematic clinical diagnosis according to a clinical interview schedule. The model is illustrated by application to multiple symptoms of depression in a large sample of female monozygotic and dizygotic twins.

THE MODEL

We distinguish two parts of the model. The "genetic" component represents the causes of family resemblance in a hypothetical latent dimension. The genetic model is summarized in the (multivariate) frequency distribution of the latent trait in the population of families from which a sample has been drawn. The second, "psychometric" component describes the relationship between the latent dimension and the responses of individuals to the test items or the symptoms on a physician's checklist.

There is already an extensive psychometric literature on "latent trait" models for psychological test data (e.g., Lord and Novick, 1968; Bock and Lieberman, 1970; Bock and Aitkin, 1981). Mislevy (1984) used a latent trait model to test for component normal distributions in the dis-

tribution of spatial ability caused by a hypothetical sex-linked locus but did not employ family data in the analysis.

There are several possible models for the relationship between a subject's score on a latent trait and the probability that he/she will endorse a particular item on a test. We assume that the probability that a given dichotomous item will be endorsed is a cumulative normal function of the subject's latent trait value. Thus, writing X_i for the response (zero–one) to the ith item and θ for the trait value of the subject, the probability of endorsement is

$$P_i(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_i(\theta - b_i)} e^{-1/2x^2} dx.$$

The parameter a_i is the "discriminating power" of the item. The trait value which results in $P_i(\theta) = 0.5$ is the "item difficulty" parameter b_i. The regression of $P_i(\theta)$ on θ is steeper at b_i for items with larger a_i (see, e.g., Lord and Novick, 1968).

For a subject of given ability, θ, the likelihood of a particular vector of k zero–one responses $\mathbf{X}$ is

$$l(\mathbf{X}|\theta) = \prod_{i=1}^{i=k} P_i(\theta)^{X_i}[1 - P_i(\theta)]^{1-X_i}.$$

In many psychometric applications, the distribution of the latent trait is not known a priori and the psychometrician is faced with the (formidable) task of estimating the item parameters and latent trait values for all the subjects. For many genetic applications, however, it is sufficient (although not necessarily easier) to estimate parameters of the distribution of θ in families sampled from the population. For example, if we are prepared to assume a large number of genes of infinitesimal effect (the "polygenic" model), then the distribution of θ is assumed to be normal and the covariances between relatives simply a function of the heritability and the degree of genetic relatedness. Under the so-called "mixed" model (e.g., Lalouel *et al.*, 1984) the distribution is assumed to be a mixture of normal distributions, each centered on the average trait value of individuals having a given genotype at a locus of large effect. If the major locus accounts for *all* the genetic variation, then the residual trait values are uncorrelated in families. If there are also polygenic effects, then the trait values will be correlated within families even when the effects of the major locus are controlled.

We write $\phi(\theta)$ for the frequency distribution of the latent trait in the population. The conditional likelihood of a given response vector, $\mathbf{X}$, is thus the integral

$$l(\mathbf{X}) = \int_{-\infty}^{\infty} \phi(\theta) l(\mathbf{X}|\theta) d\theta.$$

and the likelihood of a whole sample of unrelated individuals is simply the product of the individual likelihoods. The likelihood may then be maximized with respect to the item parameters and the parameters of the distribution of θ.

In genetic applications, however, unrelated individuals do not yield the information necessary to test alternative hypotheses about the genetic and environmental determinants of θ. For this purpose we require kinship data. For simplicity, we consider only two-member kinships, but the theoretical extension to large kinships is straightforward. As we shall see, however, the practical application of the model even with kinships of two individuals requires considerable computational resources so the treatment of larger families may be difficult in practice.

We let θ_1 and θ_2 denote the trait values of pairs of related individuals and the vectors **X** and **Y** be typical response vectors of the first and second individuals. The likelihood of the responses is then

$$l(\mathbf{X},\mathbf{Y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(\theta_1,\theta_2) l(\mathbf{X}|\theta_1) l(\mathbf{Y}|\theta_2) d\theta_2 d\theta_1 \ldots, \qquad (1)$$

where $\Phi(\theta_1,\theta_2)$ is the bivariate frequency distribution of the trait values in pairs of relatives of the given type. The form of $\Phi(\theta_1,\theta_2)$ will depend on the type of relationship being considered and the genetic model being tested. In applications where the latent trait is dependent on covariates such as sex and age, it is conceivable that each pair will have its own unique expected trait values so that Φ will have to be expressed separately for each pair. The likelihood, $l(\mathbf{X},\mathbf{Y})$, may be evaluated for each pair of relatives and the overall likelihood accumulated and maximized with respect to the item parameters and parameters of the genetic model.

APPLICATION TO TWIN DATA ON SYMPTOMS OF DEPRESSION

We illustrate the approach with data on 1983 pairs of adult Australian female twins [1233 monozygotic (MZ) and 750 dizygotic (DZ)]. The data comprise six items relating to depression from the 14-item Delusions–Symptoms–States Inventory (DSSI) designed by Bedford *et al.* (1976) for easy administration in epidemiological studies. Respondents are asked to rate their recent state with respect to each symptom on a four-point scale: "none," "a little," "a lot," and "unbearably." A detailed univariate genetic analysis of all 14 symptoms in both males and females has been published elsewhere (Kendler *et al.*, 1986). These authors describe the properties of the sample and the items in greater detail. The twins were not selected for any prior history of psychiatric disease, so no correction for ascertainment is required. In view of the extremely heavy computa-

Table I. Items Used in Analysis

Item	Recently (I have) . . .
1	been so miserable I have had difficulty sleeping.
2	been depressed without knowing why.
3	gone to bed not caring if I woke up.
4	been so low in spirits that I just sat.
5	the future seemed hopeless.
6	lost interest in just about everything.

tions, we have restricted our analysis to six of seven depression items and female twins. A seventh item, relating to "suicide," was omitted because the low endorsement frequency rendered it relatively uninformative. Females were chosen in preference to males because the higher frequency of symptoms in females (Kendler *et al.*, 1986) is expected to yield a more powerful statistical analysis.

The earlier analysis of the individual items by Kendler *et al.* has shown that an additive genetic model is adequate to account for nearly all the polychoric correlations between the responses of twins to the individual items and that there is no unequivocal evidence for nongenetic familial effects on liability to endorse the items. Jardine *et al.* (1984) confirm this finding for transformed scale scores for anxiety and depression. Kendler *et al.* (1987) conducted a multivariate analysis of the polychoric correlations using the method of weighted least squares. They showed that both genetic effects and individuals' unique nonfamilial environmental experiences contributed to the intercorrelation of the items. Neither of these earlier analyses, however, addressed specifically the issue of whether the inherited liability to depression was due to one gene or many. These earlier analyses suggested that the depression items could be selected as an illustration of the feasibility of exploiting latent trait theory in an attempt to resolve polygenic and monogenic inheritance of the dimension underlying the manifest symptoms of a psychiatric syndrome.

Table I gives the items on which the analysis is based and Table II summarizes the basic properties of the data. In computing these summary statistics, we included *all* female twins in the sample (including those from male–female pairs). Our subsequent genetic analysis includes only the female twin *pairs* in the sample. The phenotypic correlations between the items were computed by applying the product–moment formula to the raw responses coded 0, 1, 2, 3 in order of increasing severity. The first two eigenvalues of the correlation matrix are 3.359 and 0.671. There is strong support for a unidimensional model for the latent space. When the

Table II. Preliminary Statistics ($N = 4872$)

Item	Endorsement frequency	Correlation					Loading
		2	3	4	5	6	
1	0.242	0.416	0.406	0.441	0.457	0.451	0.698
2	0.351		0.401	0.471	0.404	0.459	0.692
3	0.073			0.445	0.525	0.563	0.747
4	0.195				0.486	0.538	0.755
5	0.178					0.588	0.778
6	0.120						0.812

latent trait is normally distributed and unidimensional, the item difficulties and loadings on the general factor of the tetrachoric inter-item correlations are sufficient to determine the parameters of the latent trait model (Lord and Novick, 1968). However, this will no longer be the case if a major gene is segregating, since the assumption of normality will be violated, or if genetic and environmental effects do not affect the phenotype through the same common underlying variable (e.g., Kendler *et al.*, 1987).

For the purpose of genetic analysis, the original items were recoded to be dichotomous by scoring all symptomatic responses as one and all nonsymptomatic responses as zero. Although analysis of the raw multicategory data is possible in theory, it is likely to be prohibitive in practice with our current computing resources.

We fitted three genetic models to the data. The "polygenic" model assumes additive genetic effects at a large number of loci and that no single gene has a major effect on liability so that $\Phi(\theta_1,\theta_2)$ is simply the standardized bivariate normal distribution. In the case of MZ twins the correlation in the latent trait is assumed to be h^2, i.e., the narrow heritability, and to be $\frac{1}{2}h^2$ in the case of DZ twins. The second model is the "major gene" model, in which we assume that all genetic effects on liability are due to a single major locus. Residual effects are assumed to be due to normally distributed chance environmental effects which are uncorrelated in twins. There is no barrier to the inclusion of shared environments if this proves necessary (e.g., Kendler *et al.*, 1986). We assume that there are two alleles at the major locus and that the heterozygote is exactly midway in expression between the two homozygotes. This assumption can also be relaxed to allow for dominance. In addition to the 12 item parameters, the polygenic model has a parameter for the heritability (h^2) and the major gene model has parameters for the frequency (p) of the allele which increases liability and for the additive deviation, d, of the increasing homozygote from the midpoint, m, of the homozy-

gotes. The third model is the "mixed" model, in which the genetic effects are partitioned into the effects of a major locus and residual polygenic effects. All three models allow for the effects of a normally distributed, random environmental variable. In all cases, the item parameters and genetic effects at the major locus are scaled so the latent trait has zero mean and unit variance. The polygenic component of the latent trait is assumed to be normally distributed. Kendler *et al.* (1986) showed that the frequency of "depressive" symptoms declined significantly with age ($\bar{X}$ = 35.54 years in our sample; range, 18–84 years). This effect is included in our models by estimating simultaneously the linear regression (z) of latent trait on age. Ideally, we should also like to fit the more general "unified mixed model" (Lalouel *et al.*, 1984) which seeks additional likelihood-ratio tests for the Mendelian inheritance of a hypothetical major locus but this task is not feasible with our current algorithm and computer resources.

NUMERICAL METHOD

The maximum-likelihood procedure was implemented in a FORTRAN program. Two main numerical issues have to be faced. The first is the integration of the bivariate expression (1) pair by pair to compute the likelihood for a given set of parameter values. The form of $l(\mathbf{X},\mathbf{Y})$ changes markedly with $(\mathbf{X},\mathbf{Y})$ from one pair to the next because of the unique response patterns generated by each pair. We have employed Gauss–Hermite quadrature, embodied in the Numerical Algorithms Group's (1982) FORTRAN subroutine D01FBF. Such methods approximate the integral by a weighted sum of function values for a specified optimal set of abscissae. Preliminary studies suggested that the approximation was very poor unless quite a large number of abscissae were used in each dimension (N = 12), necessitating 144 function evaluations to obtain the likelihood for each twin pair. Our finding is consistent with that reported by Bock and Aitkin (1981), who found, for the unidimensional case, that N = 10 gave very similar values for the likelihood of five LSAT responses of 1000 subjects to those obtained with N = 40.

The second numerical problem is that of maximizing the likelihood with respect to variation in the parameters. In our case, this was performed using the NAG subroutine E04JBF and associated subroutines which minimize a general function of several variables without requiring algebraic first and second derivatives of the function. This subroutine was used to minimize the negative log-likelihood ($-L$) over all twin pairs in the sample. Other investigators have used an EM algorithm in the unidimensional case to obtain item parameters without assuming any prior

distribution for the subjects' trait values (Bock and Aitken, 1981). To enhance the performance of the algorithm, $-L$ was divided by a scale factor, s, so that $L^* = L/s$ was in the range 0–1. The procedure was assumed to have converged when parameter estimates were stable to five significant figures. Typically, this criterion resulted in the gradients of L^* having values of the order of 10^{-7}.

RESULTS

The results of fitting the three models are summarized in Table III. The item parameters and major gene effects are expressed relative to the standardized liability scale. The polygenic component is expressed as a proportion of the *residual* (normally distributed) variation in liability. The item parameters do not change much as a function of the genetic model but have altered rather more than those obtained by Bock and Aitkin in their analysis of LSAT data. The item difficulty parameters show their expected consistency with the raw endorsement frequencies and there is broad agreement between the discriminating powers of the items and the loadings of the general factor extracted from the product–moment correlations. The heritability of the latent trait under the polygenic model is

Table III. Parameter Estimates Under Three Genetic Models for Multiple Symptoms of Depression[a]

	Polygenic		Major gene		Mixed	
Item	a_i	b_i	a_i	b_i	a_i	b_i
1	1.038	0.991	0.989	1.009	1.008	1.003
2	0.948	0.568	0.913	0.566	0.928	0.566
3	1.549	1.739	1.430	1.819	1.450	1.770
4	1.337	1.091	1.259	1.116	1.310	1.099
5	1.478	1.133	1.376	1.163	1.429	1.147
6	1.900	1.322	1.751	1.368	1.873	1.330
p	—		0.242		0.0064	
d	—		1.091		2.276	
h^2	0.510		—		0.463	
z	−0.0103		−0.0114		−0.0112	
L	−9073.16		−9074.57		−9070.69	
χ^2	4.94		7.76		—	
df	2		1		—	
P (%)	>5		<1		—	

[a] a_i = discriminating power of the ith item; b_i = item difficulty; p = frequency of increasing allele at major locus; d = additive deviation at major locus; h^2 = narrow heritability of polygenic component; z = regression of latent trait on age; L = log-likelihood; χ^2 = log-likelihood-ratio chi-square for comparison with mixed model.

estimated to be 0.51. This is remarkably close to the value Kendler *et al.* (1987) obtained when fitting a single common factor model to the cross-twin interitem polychoric correlations. The value given is the heritability of the latent trait and is expected to be greater than the heritability of scale scores derived from the data because the latter are subject to stochastic error. Jardine *et al.* (1984) report a narrow heritability of 0.37 for the raw depression scores for the same data.

Under the major gene model, the contribution of genetic effects to variation in liability is $g^2 = [1 - 2p(1 - p)]d^2 - [(2p - 1)d]^2$, from which we find $g^2 = 0.437$ in our data. Under the mixed model, the total genetic variation (major locus plus polygenic effects) accounts for 51% of the variation, with approximately 7% being explained by the hypothesized major locus and 41% by residual polygenic effects.

The main purpose of the analysis is to distinguish between two very different genetic models for the same data. Since the polygenic model and the major gene model are both special cases of the more general mixed model, we may see if either or both of the former receive significantly less support than the mixed model by computing the likelihood-ratio chi-square as twice the difference between two log-likelihoods (L) under the two models. We find that the mixed model is not supported better than the polygenic model ($\chi_2^2 = 4.94$, $0.05 < P < 0.1$) but that the mixed model is significantly better than the major gene model ($\chi_1^2 = 7.76$, $0.005 < P < 0.01$).

Generally, we anticipate the power of such tests to be low. Bock and Aitkin (1981) observe that ". . . the likelihood is . . . insensitive to the shape of the prior (distribution in the population so) . . . this approach could not be depended upon to estimate accurately the finer features of the ability distribution (e.g., coefficients of skewness and kurtosis) in practical sample sizes." However, in our case involving comparatively large samples, we have been able to show that the observations cannot be explained by a genetic model which invokes only one locus. It does not necessarily follow that the number of genes is large or that their effects are necessarily equal. Our analysis shows merely that no single gene has such outstanding effects as to be clearly distinguishable from any other.

DISCUSSION

A significant issue in genetic analysis of discontinuous traits is how to cope with multiple items which reflect, in different ways, the underlying phenotype. One approach is to devise more or less arbitrary "scales" prior to genetic analysis. This approach, however, is likely to be misleading because assumptions about the distribution of liability are con-

founded with those about the relationship between liability and the derived scores. Our approach separates the genetic and psychometric components of a model for multiple symptoms and allows us to test alternative genetic hypotheses about the underlying scale of liability under a plausible psychometric model for the relationship between the latent trait and the observed responses. Although our model does not avoid problems of scaling entirely, because we still have to assume a particular form for the item–trait regression, we avoid some of the more obvious pitfalls of applying likelihood-ratio tests for a major gene when the assumptions of normally distributed residual effects are clearly unwarranted. The approach is tractable with small kinships ($N = 2$) and simple genetic hypotheses. The incorporation of covariates into the model has been demonstrated by the simultaneous estimation of the regression of the latent trait on age (z). Allowing for the interaction of such covariates with the effects of a major locus presents little further practical difficulty (Eaves, 1984). The model recognizes that different symptoms may need to be given different weight in genetic analysis and allows the genetic data to decide how symptoms should be weighted to yield the best index of the latent genotype. It is important to note that this model, in common with many others which try to use multiple items or thresholds to define a scale of liability for genetic analysis, makes the strong assumption that the underlying liability is unidimensional and that genetic and environmental effects all contribute to the same dimension of liability. It is quite likely that failure of this assumption could also lead to spurious support for a major locus. A detailed multivariate analysis of the polychoric correlations for the anxiety and depression symptoms (Kendler *et al.*, 1987) suggests that genetic and environmental effects do not operate on the same underlying phenotype so our model is, at best, an approximation.

When we applied the model to multiple symptoms of depression we found no statistical evidence for a major gene, even though there was strong support for a genetic component. As long ago as the 1950s (e.g., Eysenck, 1952), it was argued that liability to psychiatric disorder might be under the same form of genetic control as variation in normal personality. Our analysis suggests that variation in the liability underlying twins' responses to set of symptoms of depression has exactly the same type of genetic control (more than one locus, additive genetic effects, intermediate heritability) as has been found repeatedly for normal variation in neuroticism. The case is strengthened still further by the recent demonstration by Jardine *et al.* (1984) of a very high genetic correlation between the neuroticism scale of the Eysenck Personality Questionnaire and the anxiety and depression scales of the DSSI.

In a more general context, our analysis illustrates an approach which can be used to generalize the segregation analysis of diseases to the case where systematic data are gathered on multiple symptoms. The statistical power of the analysis might be enhanced if the analysis were applied to data on more extended kinships ascertained through probands with extreme responses. However, the correction for ascertainment and the computational demands of the current algorithm are formidable.

Practical limitations notwithstanding, many theoretical aspects of the model are applicable to attempts to analyze genetic linkage between marker polymorphisms and diseases which are usually diagnosed on the basis of multiple symptoms. Some of the practical problems we encountered, especially the amount of computer time needed to estimate the parameters of the mixed model, may disappear with the next generation of computers. The limiting factor will be not what is technically feasible, but what can be achieved statistically with realistic sample sizes.

REFERENCES

Bedford, A., Foulds, G. A., and Sheffield, B. B. (1976). A new personal disturbance scale (DSSI/SAD). *Br. J. Soc. Clin. Psychol.* **15**:387–394.

Bock, R. D., and Aitkin, M. (1981). Marginal maximum-likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46**:443–459.

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

Cloninger, C. R., Martin, R. L., Guze, S. B., and Clayton, P. J. (1985). Diagnosis and prognosis in schizophrenia. *Arch. Gen. Psychiat.* **42**:15–25.

Eaves, L. J. (1983). Errors of inference in the detection of major gene effects on psychological test scores. *Am. J. Hum. Genet.* **35**:1179–1189.

Eaves, L. J. (1984). The resolution of genotype x environment interaction in segregation analysis of nuclear families. *Genet. Epidemiol.* **1**:215–228.

Eysenck, H. J. (1952). *The Scientific Study of Personality*, Routledge and Kegan Paul, London.

Jardine, R., Martin, N. G., and Henderson, A. S. (1984). Genetic covariation between neuroticism and the symptoms of anxiety and depression. *Genet. Epidemiol.* **1**:89–107.

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

Kendler, K. S., Heath, A. C., Martin, N. G., and Eaves, L. J. (1987). Anxiety and depression: Same genes, different environments? *Arch. Gen. Psychiat.* **44**:451–457.

Lalouel, J. M., Rao, D. C., Morton, N. E., and Elston, R. C. (1983). A unified model for complex segregation analysis. *Am. J. Hum. Genet.* **35**:816–826.

Lord, F. M., and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading, Mass.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika* **49**:359–381.

Numerical Algorithms Group (1982). *FORTRAN Subroutine Library. Mark 9*, NAG, Oxford.

Edited by C. R. Cloninger