Genetic Epidemiology 1:215–228 (1984)

The Resolution of Genotype × Environment Interaction in Segregation Analysis of Nuclear Families

Lindon J. Eaves

Department of Human Genetics, Medical College of Virginia, Richmond

A model is presented for the effects of one or two loci, a measured index of the environment and genotype × environment (G×E) interaction of risk for a discontinuous trait. Initial properties of the model are explored for the single locus case, with and without the effects of environment and G×E interaction. Seven data sets were simulated, each comprising 500 nuclear families on whom an environmental index has been measured. Maximum-likelihood estimation procedures were used to obtain parameter estimates under seven models for each data set. Likelihood ratio tests were constructed, and in all cases it was possible to identify the "correct" model for the simulated data. The matrices of information realized showed that the parameters could be estimated with acceptable precision and that the effects of genes, environment, and G×E interaction could be resolved in the simulated populations. The effects on conventional segregation analysis of ignoring the environment and G×E are considered.

Key words: genotype × environment interaction, segregation analysis, simulation, nuclear family, likelihood, environmental index, nonadditive

INTRODUCTION

Most models for the effects of genes and environment on human variation assume that the effects of genes and environment are additive. They assume there is no genotype × environment (G×E) interaction. Recent advances in the genetic and epidemiological study of common disease lead to a pressing need for tractable models of G×E interaction that can be applied to human data.

Family studies of hypertension, for example, have shown almost beyond doubt that there is a substantial genetic component to the disease [Havlik and Feinleib, 1982]. On the other hand, experimental and epidemiological studies suggest that

Received for publication May 16, 1984; revision received June 14, 1984.

Address reprint requests to Dr. L.J. Eaves, Department of Human Genetics, Medical College of Virginia, PO Box 33, MCV Station, Richmond, VA 23298.

dietary sodium may be an important environmental factor [Luft and Weinberger, 1982]. More recently, it has been argued that only certain genotypes may be especially sensitive to dietary sodium [Kawasaki et al, 1978]. Thus, there is already a growing awareness of the importance of G×E in cardiovascular disorder.

Similar considerations appear in the area of psychiatric disorders. The so-called "diathesis-stress" model is commonly used as an heuristic device in exploring the etiology of such diseases [Gottesman and Shields, 1973]. The model recognizes that some individuals are especially predisposed to respond more markedly to their environment. Two disorders for which a G×E model may be especially important are depression and criminality. Family and adoption studies have shown that depression has a significant genetic component [Gershon et al, 1976]. Epidemiological studies have identified specific environmental stresses that increase the risk of depression [Brown and Harris, 1978]. Indeed, the latter authors argue that certain individuals may be especially prone to environmental stress but, so far, no general strategy has evolved for the analysis of genetic effects on discontinuous traits with data on measured aspects of the environment.

Cloninger et al [1982] have shown how Scandanavian adoption data relating to petty criminality permit resolution of social and genetic effects on behavior. Indeed, the probability that an individual will develop a particular behavior pattern is a nonadditive function of his genotype (assessed through the phenotype of the biological parent) and his family environment (measured by the phenotype of his foster parent). Thus, on the scale of "probability," genes and environment interact in the etiology of criminality.

In purely statistical terms, the magnitude of G×E interaction is assessed by the extent to which the average performance of the genotype and the average effect of a given environment fail to predict the responses of individual combinations of genotypes and environments. The biological and genetic importance of G×E, however, can only be appreciated by examining the extensive literature of G×E in species other than man. A wealth of experimental studies have shown that G×E, though it can be described in statistical terms, is better conceived as the genetic control of sensitivity to the environment [Mather and Jinks, 1982]. Such experimental studies in fungi, higher plants, Drosophila spp, and mammals have shown beyond doubt that G×E interaction is a fundamental property of many genetic systems. More important, however, are the demonstrations that: 1) the genes that control sensitivity to the environment are often quite distinct from those that determine the average response of the individual over all environments, 2) the genes responsible for sensitivity to the environment have their own additive and dominance relationships, which may be quite different from those shown by genes that affect the average response, and 3) different genes are responsible for controlling sensitivity to different environmental factors.

In a random sample of 82 inbred lines of Nicotiana rustica, Perkins and Jinks [1971] showed that average performance and sensitivity to the environment are under separate genetic control. A chromosome assay of sternopleural chaeta number in the inbred lines Wellington and Samarkand in D melanogaster [Caligari and Mather, 1975; Mather and Caligari, 1976] showed that the genes responsible for average performance were located chiefly on chromosome III, whereas those for sensitivity to temperature were mainly on chromosome II.

Jinks and Connelly [1975] showed, in the fungus Schizophyllum commune, that the direction of selection, and the quality of the environment in which selection

occurred, had predictable consequences for the genetic control of environmental sensitivity. A more recent study of Nicotiana rustica by Jinks and Pooni [1983] confirmed these earlier experiments on a different species.

Early studies of G×E by Bucio-Alanis [1966] and Bucio-Alanis and Hill [1966] showed how the additive and dominance properties of genes responsible for sensitivity to the environment may be estimated in the same way as effects on average performance. Perkins and Jinks [1971] demonstrated how epistatic interactions could be detected between loci affecting response to the environment. Equally significant are demonstrations that the additive and dominance properties of genes controlling sensitivity and mean performance may differ, as is the case in Mather and Jinks' [1982] analysis of Powers' [1941] data from the cross Danmark × Johannisfeuer in the tomato.

Some genes controlling response to the environment may have effects that are generalized over a range of different environmental agents. The effects of other genes may be specific to individual environments. Thus, for example, Perkins and Jinks [1971] concluded in their analysis of the performance of the lines 2 and 42 of N rustica in a wide range of controlled and uncontrolled environments that there was "considerable specificity in the reactions of the genotypes to the different kinds of environmental variation." Thus, we must consider the possibility that different genes control responses to quite specific features of the environment.

All these studies point to the fact that G×E interaction is a significant component of any genetic system, that G×E has its own genetic properties distinct from those of average performance, and that G×E may be influenced quite independently by natural and artificial selection.

In spite of the evidence for G×E in other organisms, its effects have largely been ignored in human genetics. There may be two reasons for this neglect of G×E. In the first place, most of the tractable models for human variation are additive. Path analytic methods, for example, are virtually useless for deriving nonadditive contributions to family resemblance. As a result, the contribution of G×E has been relegated to residual effects for which no single powerful test is available. The second reason for the lack of interest shown by human geneticists in G×E is the fact that most approaches to the analysis of G×E in man have assumed that the environment cannot be measured directly but its effects only inferred from correlations between relatives.

Jinks and Fulker [1970] suggested one test of G×E in man that detects correlations between the genes responsible for average effect and those creating environmental sensitivity. They suggested examining the relationship between the means and standard deviations of MZ twin pairs. For twins reared apart, this would provide a test of G×E involving all postnatal environmental effects. For twins reared together, the test would only detect interactions betweeen genotype and environmental effects unique to individuals within the family. A related approach, suggested by Eaves and Eysenck [1976] is to examine the relationshp between environmental measures made on DZ twin pairs (or sibships) and the within sibship variance. This test should detect interaction between the measured environment and genetic differences within sibships. The problem with both the latter test and that devised for MZ twins reared together is that they are not specific for G×E but may also detect interaction between purely environmental factors within and between families. Furthermore, they do not help us discriminate between genes that control average performance and those that control sensitivity to the environment.

In this paper we present a model for the interaction of genetic and environmental effects that may be used in the analysis of dichotomous traits and yields, as special cases, most of the simpler models for genetic effects employed in conventional segregation analysis. Some of the features of the model are illustrated by the analysis of simulated data on nuclear families in which members are assigned to either "affected" or "unaffected" status and measured on a single environmental factor hypothesized to contribute to disease liability. There is no theoretical barrier to the incorporation of multivariate environmental indices into the model. We explore both additive and nonadditive models for the effects of genes and environment on disease liability and show how they can be resolved by a more general form of segregation analysis.

MODEL

The model is used to predict the probability of a disorder, "risk" (R), as a function of genetic and environmental components of liability, L. Since $0<R<1$, it is unlikely that genetic effects, or environmental effects, will contribute additively to risk unless the variation in liability is small. However, on the continuous scale of liability, $-\infty<L<+\infty$, the effects of genes may be additive, nonadditive, or both. Our model assumes that liability to the disease can be continuous (though might only be due to one or two genes) but that the risk to the disease, R, can be expressed as a function of liability, L, thus:

$$R = 1/[1 + \exp(-L)].$$

This is the "logistic" function, which has proved convenient in the prediction of discontinuous variables from continuous measures [eg, Kleinbaum et al, 1982] and is used widely in epidemiological studies. Most logistic regressions assume that the risk can be expressed as a function of measured variables. In our application, however, we assume that only the environment can be measured directly and allow for the effects of a latent component due to one or two genes that may either effect liability directly and equally in all environments (ie, in the absence of G×E interaction) or may effect the sensitivity to the environment in a manner comparable to that found in many animal and plant studies of environmental and genetic effects. In either case, whether or not the effects on L are additive, they could only contribute additively to risk over a very small range.

In the current version of the model it is assumed that one or two genes affect liability and that the relevant environment can adequately be summarized by a single environmental index, E. We assume two alleles at each locus and let p_a and p_b be the frequencies of the increasing alleles at loci A and B, respectively. Following the conventions of biometrical genetics [Mather and Jinks, 1982], which are more flexible than those often employed in human genetics, we let A and B denote the increasing alleles and a and b be the decreasing alleles, regardless of the dominance relationships among the alleles. We let e_j denote the jth level of the environment.

For the liability of the ith genotype in the jth environment we write:

$$L=g_i+b_ie_j,$$

where g_i is the average response of the ith genotype and b_i is the sensitivity of the ith genotype to the measured environment. The model thus assumes that liability is a

linear function of the environment in a given genotype. If b_i is zero for every genotype, there is no regression of liability on the measured environment. If the b_i are the same for all genotypes, there are environmental effects but no G×E interaction. If the b_i vary among genotypes there is G×E interaction because different genotypes respond differently to the indexed environment. The model, as we have written it, thus assumes linearity of regression of liability on environment. This is not a necessary restriction but one which, in practice, has been found effective [eg, Jinks and Connelly, 1975].

In a model involving only one or two loci we may devise parameters to express the average effects, g_i, of the loci on liability and their effects on sensitivity to the environment, b_i. We define *m* as the midpoint in liability between the aabb homozygote and the AABB homozygote in the average environment. Then for each locus we may define additive deviations from *m*: d_a and d_b. Similarly, the deviation of the heterozygote at each locus from *m* can be specified: h_a and h_b. This parameterization of the gene effects is that employed by Mather and Jinks [1982] and captures both the usual cases encountered in human genetics and the more general cases explored in the genetic analysis of complex variables in other species. Thus, if the heterozygous effect, *h*, is zero at a given locus, the heterozygote is exactly intermediate in liability between the two homozygotes. This corresponds to the classical case of "codominant" inheritance. If $h = d$ at a locus, the allele that increases liability is dominant. This corresponds to the classical case of "dominance" for the disease. If $h = -d$, then the allele that decreases liability is dominant. This case corresponds to the classical "recessivity" for the disease state. Various degrees of "partial dominance" (or recessivity) are captured by intermediate values of $-d < h < d$.

Since the model allows for the effects of two loci on liability, we may also incorporate, in theory at least, the epistatic interactions between loci. Following Mather and Jinks [1982] we recognize that four parameters are required for a complete specification of digenic interactions:

i_{ab} is the interaction between the homozygotes AA and BB;

j_{ab} is the interaction between the AA homozygote and the Bb heterozygote;

j_{ba} is the interaction between the Aa heterozygote and the BB homozygote;

l_{ab} is the interaction between the heterozygotes Aa and Bb.

Mather and Jinks show how different constraints on the four interaction parameters correspond to the classical instances of digenic interaction including classical duplicate and complementary gene interaction. The model, however, recognizes that the classical forms of epistasis are merely special cases of a more general model for digenic interactions. The resolution of epistatic effects is likely to be difficult in man but the general model leads to a better understanding of the arbitrariness that accompanies the specification of epistatic effects in many attempts at segregation analysis under the two-locus model.

The effects of each genotype on sensitivity to the environment may be specified similarly in terms of additive and dominance effects. Thus, we define:

g_m = the mean sensitivity of the AABB and aabb homozygotes to the measured environment;

g_{da} = the deviation in sensitivity of the AA homozygote from the average sensitivity;

g_{ha} = the deviation in sensitivity of the heterozygote AA from the mean sensitivity to the environment.

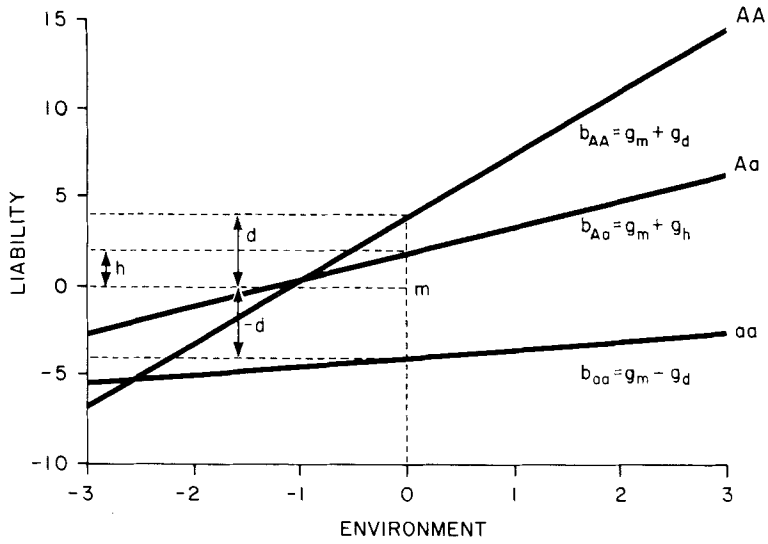

Fig. 1 The relationship between liability and environment under the genotype × environment interaction model. See text for definitions of parameters.

Similarly, we may define the additive and heterozygous effects of the B/b locus on sensitivity to the environment. If the regression of the heterozygote's liability on the environment is exactly midway between that of the two homozygotes, then there is no dominance for sensitivity to the environment. If, in addition, the two homozygotes are equally sensitive to the environment on the scale of liability and there is no genotype × environment interaction in any sense that is consistent with the concept of G×E in quantitative genetics. Theoretically, we could specify epistatic effects of the two genes on sensitivity to the environment but such sophistication is almost certain to be of academic interest only in the analysis of human differences.

The two-locus model has been specified in some detail because it embodies, as special cases, a number of important alternative hypotheses about the additive and nonadditive effects of genes and environment on disease liability, which have been all but ignored in the analysis of segregation.

For example, the animal and plant studies described above give reason to specify a model in which different genes affect average response and sensitivity to the environment. The model allows for this possibility when the regression parameters g_d and g_h are set equal to zero at the first locus (thus allowing no genetic effect of the first locus on sensitivity) and the average effect d_b and h_b are equal to zero at the second locus. Clearly, there are many alternative models that are special cases of the general two-locus model for G×E we have described. It remains to be seen how very subtle alternatives can be distinguished.

The interpretation of the parameters for a single locus is illustrated in Figure 1, which shows the regression of the liability of each genotype on the environmental index as a function of the additive and dominance components of average effect and environmental sensitivity.

TABLE I. Expected Liabilities of Genotypes at Two Loci as Function of Additive and Nonadditive Genetic Effects, Measure Environmental Effects (e_j), and Their Interaction With Genotype (G×E)

		Genetic effects								Environment + G×E[a]				
		Additive		Dominant		Epistatic								
Genotype	m	d_a	d_b	h_a	h_b	i_{ab}	j_{ab}	j_{ba}	l_{ab}	g_m	g_{da}	g_{da}	g_{ha}	g_{hb}
AABB	1	1	1	0	0	1	0	0	0	1	1	1	0	0
AABb	1	1	0	0	1	0	1	0	0	1	1	0	0	1
AAbb	1	1	−1	0	0	−1	0	0	0	1	1	−1	0	0
AaBB	1	0	1	1	0	0	0	1	0	1	0	1	1	0
AaBb	1	0	0	1	1	0	0	0	1	1	0	0	1	1
Aabb	1	0	−1	1	0	0	0	−1	0	1	0	−1	1	0
aaBB	1	−1	1	0	0	−1	0	0	0	1	−1	1	0	0
aaBb	1	−1	0	0	1	0	−1	0	0	1	−1	0	0	1
aabb	1	−1	−1	0	0	1	0	0	0	1	−1	−1	0	0

[a]The coefficients of the environmental and G×E component are multiplied by the environmental index measure, e_j, to give the expected liability in a given environment.

In Table I we give the expected liability of each genotype in a given environment in terms of the parameters of the two-locus model.

Further grasp of the model may be obtained from Figure 2, which gives the risk of the disorder as a function of genotype and environment for a special case in which the A/a gene has effects both on average liability and sensitivity to the environment.

The three curves describe, for each genotype at the A locus, the probability of being affected as a function of the independent environmental index, E. The curves, therefore, show the variation in penetrance of the three genotypes as a function of the environmental index. In drawing the curves we have assumed the following parameters are nonzero: $d_a = 2$, $g_m = 2$, $g_{da} = 2$. Thus, the heterozgote is assumed to be intermediate in penetrance and liability in the average environment ($E = 0$). The AA homozygote is assumed to increase average effect and show increasing liability as the environmental index increases, whereas the combination of a positive average regression ($g_m = 2$) with the effect of the decreasing homozygote on sensitivity to the environment ($-g_{da} = -2$) makes the penetrance and liability of the aa homozygote constant in all environments. This particular set of parameter values predicts a reversal of the ranking of the three genotypes, and their dominance relationships, as the environment increases beyond $E = -1$. Such crossover interactions, though not ubiquitous, are not unknown in careful studies of G×E interaction.

MAXIMUM-LIKELIHOOD ESTIMATION

The above model has been implemented in a FORTRAN program for the segregation analysis of nuclear families, which employs copyright software from the Numerical Algorithms Group's [1982] FORTRAN library of numerical analysis programs. The program employs the NAG subroutines E04HBF and E04JAF for minimization of a general function of many variables subject to linear and nonlinear constraints. The main difference between segregation analysis under the more classical models and under the model described here is that each individual in the sample

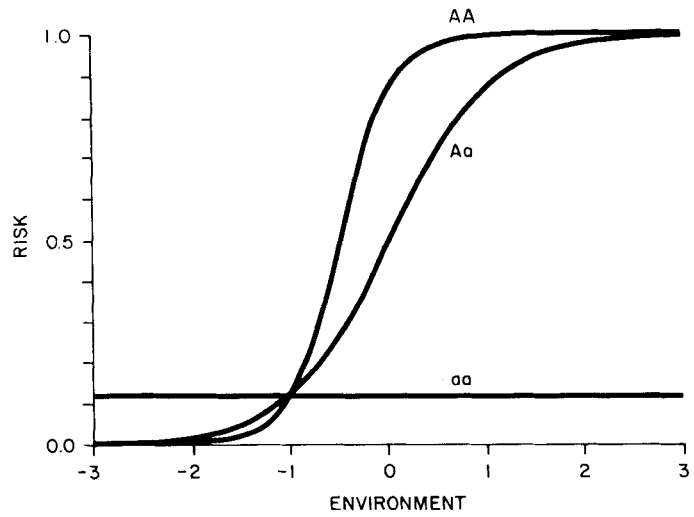

Fig. 2 The relationship between risk and measured environment for the three genotypes for one of the simulated examples. Parameter values assumed: $m = 0$, $d_a = 2$, $g_m = 2$, $g_d = 2$, $h = g_h = 0$.

has a unique probability of being affected that is a function of genotype at the two loci and the value of the environmental index. Thus, for a given mating type, the probability that an offspring will be affected is not the same for all children but varies with the environment. The program computes and maximizes the likelihood over all families, given the measured values of the environmental index and permits some or all of the parameters of the model to be free or fixed. Sampling may be random, or the ascertainment probability, π , may be fixed or estimated as desired.

The method of maximum-likelihood may be applied to the estimation of parameters under a number of alternative hypotheses in order to compare models in which there is no regression on the environmental index with those in which the environment affects liability and models that assume additive effects of genes and environment on liability with those in which there are genetic effects on sensitivity to the environment (G×E interaction). The matrix of information realized at the final solution is computed for free parameters using a numerical procedure suggested by Davis and Polonsky [1965] and inverted to provide approximate variances and covariances for the parameter estimates.

SIMULATION

The segregation analysis program was implemented in the VCU Amdahl computer and employed in the analysis of seven sets of nuclear family data simulated on the Department of Human Genetics PDP 11-44 computer using a FORTRAN 77 program. Each set comprised 500 families of two parents and four children. In this simulation study it was assumed that the families were sampled at random from a population in which a single common gene was segregating ($p_a = 0.3$) and that there was a single measure of the environment, which was N[0,1]. In these simulations the environment was assumed to be distributed randomly with respect to family members.

The seven data sets were simulated to represent all the main combinations of genetic, environmental, and G×E effects possible in the single locus model though clearly there are many more possibilities than those explored in this basic treatment. It was assumed throughout that the gene effects on average liability and sensitivity were additive, ie, the heterozygote is intermediate in average liability and in sensitivity to the environment.

The seven data sets were simulated as follows: Set 1: average genetic effect only, no environmental effect or G×E ($g_m = 0$, $g_{da} = 0$); Set 2: environmental regression only, no genetic effects or G×E; Set 3: gene affects sensitivity to environment only ("G×E only") and no average effect; Set 4: average genetic effect, environmental effect, no G×E; Set 5: average genetic effect, G×E due to same gene; $g_m = 0$; Set 6: average regression on environment and G×E but no average genetic effect on liability; Set 7: average genetic effect, average environmental effect, and G×E interaction due to same locus.

The computer program for maximum-likelihood estimation was employed to analyze each of the seven data sets in turn. The seven models described above were fitted to each of the seven data sets. Parameter estimates were obtained for each model. The matrices of information realized were computed for each data set only for the model employed in simulating each particular set.

RESULTS

Table II summarizes the parameter estimates for all seven models fitted to the seven data sets.

The inverse of the matrix of information realized is given only for the seventh data set, which was simulated with genetic, environmental, and G×E interaction effects in Table III.

In each case, model 7 is regarded as the "full model" because it includes genes, environment, and G×E. This model should give the highest likelihood (L) or the smallest value of −L. The same model should not fit significantly better than any submodel in which redundant parameters are deleted. Deletion of parameters that are making a "real" contribution to variation and family resemblance, however, should lead to a significant reduction in likelihood. We assume that any change in log-likelihood greater than 1.92 associated with removal of a single parameter indicates the significance of that effect at the 5% level.

For every data set, we find that these criteria lead us to the correct decision about the mode of inheritance. For example, in the first data set, which was simulated on the assumption that liability (though not "risk") was due entirely to the effects of a single additive locus without any environment effect or G×E, the full model does not offer a significant improvement on the simple model that assumes exactly what was specified in simulating the data. The likelihood under the full model is −1891.11, when the effects of the environment and G×E are eliminated from the model, support only increases by 2.08 units for 2 df. By contrast, all models in which the main effect of the gene is excluded are far less well supported, typically by over 50 units.

The same trend is found in all cases. The original model is recovered with great reliability for the parameter values assumed in the simulations. Some apparent anomalies can be explained. For example, in the second data set, for which only environmental effects are responsible for the observed trait, we note that the model

TABLE II. Results of Fitting Models for Genetic, Environmental, and G×E Interaction to Seven Simulated Data Sets

Data set	Model	Parameters					−L
		p_a	m	d_a	g_m	g_d	
1	Given	0.3	0.	2.0	0.	0.	—
	1	0.269	0.206	2.109	0[a]	0[a]	1893.19
	2	0[a]	−0.618	0[a]	0.021	0[a]	1942.81
	3	1[b]	−0.618	0[a]	0[a]	0.021	1942.81
	4	0.272	0.189	2.104	0.029	0[a]	1893.04
	5	0.427	−0.545	2.047	0[a]	0.503	1891.99
	6	0.010	−0.631	0[a]	9.985	9.997	1942.32
	7	0.299	−0.066	2.162	0.162	0.513	1891.11
2	Given	0.3	0.	0.	2.0	0.	—
	1	0.485	0.023	−0.000	0[a]	0[a]	2079.25
	2	0[a]	0.045	0[a]	2.023	0[a]	1390.30
	3	1[b]	0.045	0[a]	0[a]	2.023	1390.30
	4	0.440	0.045	0.000	2.023	0[a]	1390.30
	5	1[b]	−0.549	0.594	0[a]	2.023	1390.30
	6	0.226	0.045	0[a]	2.023	0.000	1390.30
	7	0.964	−3.647	3.762	10.000[b]	−8.052	1389.03
3	Given	0.3	0.	0.	0.	2.0	—
	1	0.489	0.052	0.000	0[a]	0[a]	2078.43
	2	0[a]	0.045	0[a]	−0.498	0[a]	1993.59
	3	0.280	0.046	0[a]	0[a]	1.610	1973.26
	4	0.529	0.045	−0.005	−0.498	0[a]	1993.59
	5	0.281	0.067	0.054	0[a]	1.616	1973.22
	6	0.261	0.046	0[a]	0.080	1.637	1973.23
	7	0.264	0.067	0.049	0.069	1.638	1973.20
4	Given	0.3	0.	2.0	2.0	0.	—
	1	0.370	−0.187	0.961	0[a]	0[a]	2012.34
	2	1[a]	−0.595	0[a]	1.622	0[a]	1495.60
	3	1[b]	−0.595	0[a]	0[a]	1.622	1495.60
	4	0.283	−0.008	1.566	1.935	0[a]	1481.84
	5	0.983	−10.	9.475	0[a]	1.682	1493.21
	6	0[b]	−0.000[b]	0[a]	1.641	0.019	1495.60
	7	0.610	−1.151	1.627	2.059	−0.312	1481.77
5	Given	0.3	0.	2.0	0.	2.0	—
	1	0.475	−0.501	1.213	0[a]	0[a]	1981.15
	2	1[b]	−0.494	0[a]	−0.361	0[a]	1950.39
	3	0.324	−0.583	0[a]	0[a]	1.330	1933.43
	4	0.057	2.830	3.666	−0.490	0[a]	1928.78
	5	0.308	−0.026	2.482	0[a]	2.438	1876.81
	6	0.149	−0.626	0[a]	1.068	2.107	1928.66
	7	0.289	0.056	2.402	0.089	2.373	1876.69
6	Given	0.3	0.	0.	2.0	2.0	—
	1	0.486	0.069	−0.000	0[a]	0[a]	2077.64
	2	1[b]	0.063	0[a]	0.705	0[a]	1928.61
	3	0.777	0.064	0[a]	0[a]	1.729	1913.53
	4	0.585	0.063	−0.000	0.705	0[a]	1928.61
	5	0.776	0.235	−0.345	0[a]	1.735	1912.05
	6	0.263	0.066	0[a]	2.177	2.098	1909.58
	7	0.261	−0.141	−0.324	2.211	2.132	1908.17

(continued)

TABLE II. Results of Fitting Models for Genetic Environmental, and G×E Interaction to Seven Simulated Data Sets (continued)

Data set	Model	Parameters					−L
		p_a	m	d_a	g_m	g_d	
7	Given	0.3	0.	2.0	2.0	2.0	—
	1	0.122	0.685	2.009	0[a]	0[a]	1851.23
	2	1[b]	−0.834	0[a]	0.784	0[a]	1715.43
	3	0.848	−0.869	0[a]	0[a]	1.294	1710.01
	4	0.815	−10.000[b]	9.943	1.271	0[a]	1663.21
	5	0.714	−2.101	2.382	0[a]	2.100	1641.90
	6	0.119	−0.939	0[a]	4.984	4.547	1694.77
	7	0.273	0.098	2.003	2.112	2.049	1639.35

[a]Parameter fixed ex hypothesi.
[b]Parameter on boundary.

that assumes only G×E (instead of environmental effects) gives an identical likelihood to the environmental model but that the gene frequency becomes fixed at the upper bound of unity, making the model equivalent to the pure environmental model. The absence of polymorphism for sensitivity to the environment corresponds to the conventional environmental model. Gene frequencies carry no information in cases where the additive genetic effect is close to zero.

The recovery of the true parameters is easiest in the first three data sets in which the main effects on liability are assumed to operate one at a time. In the fourth and fifth data sets, which reflect a gene of additive effect either with environmental effects (Set 4) or G×E (Set 5) the true causes of inheritance are identified with these large samples without difficulty. In the sixth data set, for which environmental effects and G×E are present without an average additive deviation, the "correct" model is only about 3 units better supported than are models in which the environmental effect is deleted or replaced by a (nonsignificant) additive average genetic effect. Although such a difference is significant in samples of this size, the power of the test is likely to be low in smaller studies. Similar conclusions follow from the last data set. Although the full model is the best supported and the parameter estimates are very close to those assumed in simulating the data, we find that the model that sets to zero the midhomozygous regression on the environment is only likely to be significantly worse in samples of comparable magnitude to those we have employed.

Examination of the covariances and standard errors given in Table II reveals that the standard errors of the parameter estimates are acceptably small, given these sample sizes, and confirm, for the case of the most complex model, that the estimated values of the parameters come acceptably close to those provided as population values to the simulation program. Thus, in this case, the parameter estimates are not seriously biased when the right model is fitted. This may not be true in general for more extreme parameter values, for example, for very small gene frequencies. The large correlations between some parameters may detract from the power of the design for resolving G×E in some circumstances.

By contrast, when complex effects are ignored, serious biases may occur in parameter estimates. For example, if the data sets generated are subjected to analyses that ignore the effects of environment and genotype × environment interaction, errors

TABLE III. The Inverse of the Matrix of Information Realized at the ML Parameter Values for a Simulated Data Set Involving Genetic, Environmental, and G×E Effects*

Parameter	p_a	m	d_a	g_m	g_{da}
p_a	0.00142	−0.00767	−0.00208	−0.00907	−0.00419
m	−0.867	0.05508	0.02717	0.06120	0.03484
d_a	−0.289	0.606	0.03653	0.02565	0.02216
g_m	−0.778	0.843	0.434	0.09562	0.06865
g_{da}	−0.417	0.556	0.435	0.832	0.07119
Estimate	0.273	0.098	2.003	2.112	2.049
se	0.038	0.235	0.191	0.309	0.267

*Variances on diagonal, covariances in upper triangle, correlations between estimates in lower triangle.

of inference can occur. The first model assumes that variation in liability is due only to a single gene. The heteroygote and midhomozygote correspond to the "zero" point on the scale of liability, given the parameter values assumed in simulating the data. The heterozygote thus has, in an average environment, a 50% chance of manifesting the affected phenotype. Since the two homozygotes deviate by 2 liability units in either direction, the probability that the aa homozygote will be affected is 0.1192 and the AA homozygote has a penetrance of 0.8808. The simple additive genetic model, therefore, allows for purely stochastic error in the translation of liability into risk. By altering the additive and dominance deviations any set of "penetrance" values may be captured by the logistic model in its simplest form. However, the general model allows for more complex factors to affect penetrance, including environmental effects and G×E. It is instructive to see how such effects contribute to biases when models are fitted that assume they are absent. If we simply consider the seven data sets described here and examine what happens when we fit just the conventional single-gene model, we find that the first data set indeed is consistent with the single-gene model, with due allowance for stochastic effects between liability and risk, and that the parameter estimates are close to those used to simulate the data.

In the second and third data sets there is also no problem, because the effects of environment and G×E interaction eradicate all evidence of an average effect of genotype on liability. In these cases, even when the additive genetic effect is allowed to take its own value, it rapidly approaches zero. The same is true for the sixth data set, in which the environment and genetic sensitivity alone contribute to liability. There is no chance that a major gene will be wrongly identified in these cases because, when the environment is random as we have assumed, there is no family aggregation of the phenotype even when there is genetic sensitivity to the environment. It should be noted, however, that the absence of an average effect of the gene in the sixth data set does not mean that genetic studies are uninformative. Rather, classical genetic studies would miss the effects that are expressed only in the right environment.

The problems of bias and mistaken inference become acute when the effects of genes operate against a background of environmental effects or G×E that cannot simply be written off as differences in "penetrance" between the three genotypes. In this respect data sets 4, 5, and 7 are especially instructive. Fitting the simplest genetic model to data set 4 gives a gene frequency estimate of 0.37 compared with the true value of 0.3, and an additive genetic deviation (average effect) of 0.961 compared

with the true value of 2.0. The reason for the difference in average effect lies in the fact that the model is forced to assign all the effects of the environmental variable to differences in penetrance. With the fifth data set, which reflects the effects of the same gene on average liability and sensitivity to the environment, we find that the estimated gene frequency is seriously biased and, if we ignored G×E interaction, our interpretation of the data would be seriously impoverished. In the last data set, in which there are genetic, environmental, and G×E interaction effects, misspecification of the model leads to wild variation in the estimates of the gene frequency. If the conventional model is assumed, the estimated gene frequency is 0.122 (Set 1). If environmental effects are ignored but G×E specified in addition to the average effects (Set 5) the gene frequency estimate is 0.714. Allowing for environmental effects but ignoring G×E gives an estimate of 0.119 (Set 6).

DISCUSSION

These simulations are not exhaustive. Several important questions remain unanswered but, nevertheless, are well within the scope of the model and method described here.

We need further studies to determine whether or not: 1) the inclusion of heterozygous effects in the model ("dominance or recessivity") detract from the resolution of G×E from other genetic effects; 2) the average effects of one gene and the effects of a second on sensitivity can be resolved from the effects of one gene on both average liability and sensitivity; 3) the power of the method is affected by correlations between the environments of family members; and 4) the power of the method is affected by the problems of ascertainment associated with less frequent traits.

These issues notwithstanding, however, the simulations begin to investigate the extent to which segregation analysis may resolve more subtle effects than those considered in most treatments so far.

Recent criticism of methods commonly used in genetic epidemiology [eg, Karlin et al, 1981] has focused on assumptions commonly made in analyzing human differences. One such criticism is that the effects of genotype × environment interaction are ignored. Unfortunately, the criticism may be given undue credence in the absence of a viable approach for the statistical investigation of G×E and any firm understanding of the biological and clinical significance of such interactions. Though it is true that many human geneticists have dismissed interactions as being of secondary importance, Karlin et al do not offer any clear guidelines for the analysis or understanding of G×E interaction. The model and simulations reported in this study show how the effects of environmental heterogeneity and G×E interaction can be examined without altering the basic principles of model-fitting, which have given such a strong direction to quantitative genetic research in man and other species. Though further investigation is certainly needed, our initial simulations suggest that significant genetic information may be missed in analyses that ignore G×E when it is present. In particular, estimates of the gene frequency may be biased. With the growing awareness in epidemiology that only certain genotypes may be especially sensitive to particular environments, the basic model described here may have some value in conceptualizing and analyzing such interactions of genes and environment.

ACKNOWLEDGMENTS

This research was supported by the Commonwealth of Virginia and NIH grant GM80520-01. I thank my colleagues, Drs. A.C. Heath, K.S. Kendler, N.G. Martin, and W.E. Nance for their advice and encouragement.

REFERENCES

Brown GW, Harris T (1978): "Social Origins of Depression. A Study of Psychiatric Disorder in Women." London: Tavistock.

Bucio-Alanis L (1966): Environmental and genotype-environmental components of variability: I. Inbred lines. Heredity 21:387–397.

Bucio-Alanis L, Hill J (1966): Environmental and genotype-environmental components of variability: II Heterozygotes. Heredity 21:399–405.

Caligari PDS, Mather K (1975): Genotype-environment interaction. III. Interactions in *Drosophila melanogaster*. Proc R Soc Lond B 191:387–411.

Cloninger CR, Sigvardsson S, Bohman M, et al (1982): Predisposition to petty criminality in Swedish adoptees: II: Cross-fostering analysis of gene-environment interaction. Arch Gen Psychiatry 39:1242–1247.

Davis PF, Polonsky I (1965): Numerical interpolation, differentiation and integration. In Abramowitz M, Stegan IA (eds): "Handbook of Mathematical Functions." New York: Dover.

Eaves L J, Eysenck HJ (1976): Genotype × age interaction for neuroticism. Behav Genet 6:359–362.

Gershon ES, Bunney WE, Leckman JF, Ferdewegh M, Debauche BA (1976): The inheritance of affective disorders: A review of data and hypotheses. Behav Genet 6:227–261.

Gottesman II, Shields J (1973): Genetic theorizing and schizophrenia. Br J Psychiatry 122:15–30.

Havlik RJ, Feinleib M (1982): Epidemiology and genetics of hypertension. Hypertension 4 [Suppl III]: 121–127.

Jinks JL, Connelly V (1975): Determination of the environmental sensitivity of selection lines by the selection environment. Heredity 34:401–406.

Jinks JL, Fulker DW (1970): Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. Psychol Bull 73:311–349.

Jinks JL, Pooni HS (1983): Determination of the environmental sensitivity of selection lines of *Nicotiana rustica* by the selection environment. Heredity 49:291–294.

Karlin S. Williams PT, Carmelli D (1981): Structured exploratory data analysis (SEDA) for determining mode of inheritance of quantitative traits. I. Simulation studies on the effect of background distribution. Am J Hum Genet 33:262–281.

Kawasaki T, Delea CS, Bartter FC, Smith H (1978): The effect of high-sodium and low-sodium intakes on blood pressure and other related variables in human subjects with idiopathic hypertension. Am J Med 64:193.

Kleinbaum DG, Kupper LL, Chambless LE (1982): Logistic regression analysis of epidemiologic data: Theory and practice. Commun Statist Theor Meth 11:485–547.

Luft FC, Weinberger MH (1962): Sodium intake and essential hypertension. Hypertension 4 [Suppl III]: 14–19.

Mather K, Caligari PDS (1976): Genotype × environment interactions. IV. The effect of the background genotype. Heredity 36:41–48.

Mather K, Jinks JL (1982): "Biometrical Genetics. The Study of Continuous Variation" London: Chapman and Hall.

Numerical Algorithms Group (1982): "FORTRAN: Library Manual, mark 9, Vol 3." Oxford, England: Numerical Algorithms Group.

Perkins JM, Jinks JL (1971a): Analysis of genotype × environment interaction in triple test cross data. Heredity 26:206–209.

Perkins JM, Jinks JL (1971b): Specificity of interaction of genotypes with contrasting environments. Heredity 26:463–474.

Powers L (1941): Inheritance of quantitative characters in crosses involving two species of *Lycospersicon*. J Agr Res 63:149–174.

Edited by D.C. Rao