Am. J. Hum. Genet. 65:531–544, 1999

Testing the Robustness of the Likelihood-Ratio Test in a Variance-Component Quantitative-Trait Loci–Mapping Procedure

David B. Allison,[1] Michael C. Neale,[2] Raffaella Zannolli,[1,3] Nicholas J. Schork,[4] Christopher I. Amos,[5] and John Blangero[6]

[1]Obesity Research Center, St. Luke's/Roosevelt Hospital, Columbia University College of Physicians & Surgeons, New York; [2]Departments of Psychiatry and Psychology, Virginia Institute for Psychiatry and Behavioral Genetics, Virginia Commonwealth University, Richmond; [3]Department of Pediatrics, Policlinico LeScotte, University of Siena, Siena, Italy; [4]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, and Department of Biostatistics and Program for Population Genetics, Harvard University School of Public Health, Boston; The Jackson Laboratory, Bar Harbor, ME; [5]Department of Epidemiology, M.D. Anderson Cancer Center, Houston; and [6]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio

Summary

Detection of linkage to genes for quantitative traits remains a challenging task. Recently, variance components (VC) techniques have emerged as among the more powerful of available methods. As often implemented, such techniques require assumptions about the phenotypic distribution. Usually, multivariate normality is assumed. However, several factors may lead to markedly nonnormal phenotypic data, including (*a*) the presence of a major gene (not necessarily linked to the markers under study), (*b*) some types of gene × environment interaction, (*c*) use of a dichotomous phenotype (i.e., affected vs. unaffected), (*d*) nonnormality of the population within-genotype (residual) distribution, and (*e*) selective (extreme) sampling. Using simulation, we have investigated, for sib-pair studies, the robustness of the likelihood-ratio test for a VC quantitative-trait locus–detection procedure to violations of normality that are due to these factors. Results showed (*a*) that some types of nonnormality, such as leptokurtosis, produced type I error rates in excess of the nominal, or α, levels whereas others did not; and (*b*) that the degree of type I error–rate inflation appears to be directly related to the residual sibling correlation. Potential solutions to this problem are discussed. Investigators contemplating use of this VC procedure are encouraged to provide evidence that their trait data are normally distributed, to employ a procedure that allows for nonnormal data, or to consider implementation of permutation tests.

Received August 19, 1998; accepted for publication May 28, 1999; electronically published July 13, 1999.

Address for correspondence and reprints: Dr. David B. Allison, Obesity Research Center, St. Luke's/Roosevelt Hospital, Columbia University College of Physicians & Surgeons, 1090 Amsterdam Avenue, 14th Floor, New York, NY. E-mail: dba8@columbia.edu

0002-9297/99/6502-0029$02.00

Introduction

Obtaining sufficient statistical power to detect genes influencing complex quantitative traits (i.e., quantitative-trait loci [QTLs]) remains a daunting task (Risch and Merikangas 1996; Allison and Schork 1997). One class of techniques that has emerged as more powerful than certain alternatives is the estimation of variance components (VC) (Schork 1992, 1993; Amos 1994; Amos et al. 1996, 1997; Fulker and Cherny 1996; Wijsman and Amos 1997; Almasy and Blangero 1998; Williams and Blangero 1999). Therefore, investigators may justifiably be motivated to use these techniques. Indeed, VC techniques are beginning to be incorporated into genome scans of quantitative traits (e.g., see Comuzzie et al. 1997; Pratley et al. 1998). In the context of genetic "model-free" VC mapping in humans, one typically expresses the phenotypic variances and covariances among related individuals as a function of the (estimated) number of alleles shared identical by descent (IBD), at a genetic locus, by those relatives.

However, it is noteworthy that such techniques, as usually formulated, involve maximum-likelihood estimation and likelihood-ratio testing. Specification of the likelihood requires assumption of a particular error distribution within each IBD class. Typically, a multivariate normal distribution is assumed. If the assumed distribution is different from the true error distribution, two classes of problems may occur. First, under the null hypothesis of no linkage between the marker under study and the QTL, the type I error rate may be higher than the specified nominal level—hereafter referred to as the "nominal α level"—or, conceivably, may be lower than this level. Second, if linkage exists, the estimate of the QTL effect may be biased, and statistical power may be affected. In this report, we confine our attention to the first type of problem. We hope that future research will consider the second type of problem (also see Amos et

al. 1996; Almasy and Blangero 1998; Williams and Blangero 1999; Wang et al., in press).

There are several reasons why nonnormality might occur. First—and perhaps foremost—investigators generally undertake QTL-mapping efforts when there is strong prior evidence for a "major" gene or oligogene. Such major or oligogenes will create mixed nonnormal distributions (Matthysse et al. 1979; Schork et al. 1990*a*, 1990*b*, 1996). Although the "within-genotype" or residual distribution may be normal, this does not imply that the overall trait distribution in the population at large will be normal. In fact, one would never want to see a normally distributed phenotype for mapping purposes (see Discussion section). Of course, the presence of a major gene or oligogene implies that somewhere in the genome the null hypothesis of no oligogenic effects is false. However, it does not imply that it is false for the marker locus under study. This point was raised by Lander and Botstein (1989, Appendix A4) in the context of QTL mapping with experimental organisms.

A second possible source for nonnormality is gene × environment (G × E) interaction (Pooni and Jinks 1976). This can be the interaction either between a major gene or oligogene and the environment or between a polygenic component and the environment.

A third possibility is that a phenotype may simply be "intrinsically" nonnormal. By intrinsically nonnormal, we simply mean that it is not normally distributed, for reasons that have nothing to do with selective sampling, the mixture distributions created by major genes or oligogenes, G × E interactions, or the artificial dichotomization of a continuous trait.

Fourth, some traits may be truly dichotomous or binary, such as diagnosis of type I diabetes. In other cases, investigators either arbitrarily dichotomize a continuous distribution into "high" and "low" categories and treat the new discrete variable as a phenotype taking a value of either 0 or 1 (e.g., see Motum et al. 1993) or assume a threshold process with an underlying multivariate normal distribution (Duggirala et al. 1997). In either case, the result is a clearly nonnormal distribution.

Finally, another possible reason for nonnormality is selective (extreme) sampling. It is now well established that, under some (but not all) circumstances, selection of individuals or relative pairs that are phenotypically extreme can dramatically increase statistical power (Eaves and Meyer 1994; Risch and Zhang 1995, 1996; Allison et al. 1998). However, if the overall population distribution is normal, then the distribution among those selected on the basis of having extreme phenotypes cannot be normal. In theory, this can be managed appropriately in two ways. One of these ways involves incorporating the phenotypic scores from the individuals not selected for genotyping and using a "full likelihood" implementation, as described elsewhere (Kruglyak and Lander 1995; Eaves et al. 1996). However, this requires knowledge of the phenotypes of the unselected individuals, which may not be available. The other approach involves incorporating the selection probabilities into the calculation of the likelihood (de Andrade et al. 1997). However, this requires statistically adjusting the likelihood of the data for the ascertainment event and may reduce the power to detect linkage. Moreover, in some cases, the investigator may not explicitly know the probability of an ascertainment event (e.g., when he or she is recruiting very thin and obese subjects by placing a public advertisement that simply requests very thin and obese people to respond). Thus, when using selective sampling, an investigator employing VC procedures may be forced to ignore (or naively choose or be unaware of) the selection and, thereby, violate an assumption underlying the calculation of the test statistic.

In this report, we consider nonnormality due to these factors, as well as its effects on the type I error rate. We conduct our simulations in the context of a sibling-pair study and use the VC test described by Fulker and Cherny (1996). Given data on sib-pairs only, the test described by Fulker and Cherny is identical to that described by Amos (1994) and Almasy and Blangero (1998), when IBD likelihoods are used or when the markers are perfectly informative.

Material and Methods

Design

Our design was 2 levels of sample size × 12 distributions × 3 degrees of residual correlation. By "residual correlation" we mean the correlation between siblings' phenotypes after the genotype's effects at the QTL have been removed. For siblings, this correlation is induced by both other genetic influence on the trait and the shared-common-environment influences on the trait (Neale and Cardon 1992). A description of the levels for each factor is provided below. For each parameter set, 100,000 simulated data sets were generated under the null hypothesis of no linkage at the marker locus. Type I error rates were evaluated at putative α levels of .05, .01, .001, and .0001. We also tested the procedure and our software with data randomly sampled from a bivariate normal distribution.

Generation of phenotypic data.—For the generation of phenotypic data, we used 12 experiments. Sample histograms of the marginal distributions used in our study have been plotted in figure 1, and the population means, variances, skewness, and kurtosis coefficients are given in table 1. The values in table 1 were obtained analytically.

Normal distribution.—First, we simulated data from a bivariate normal distribution under the null hypoth-

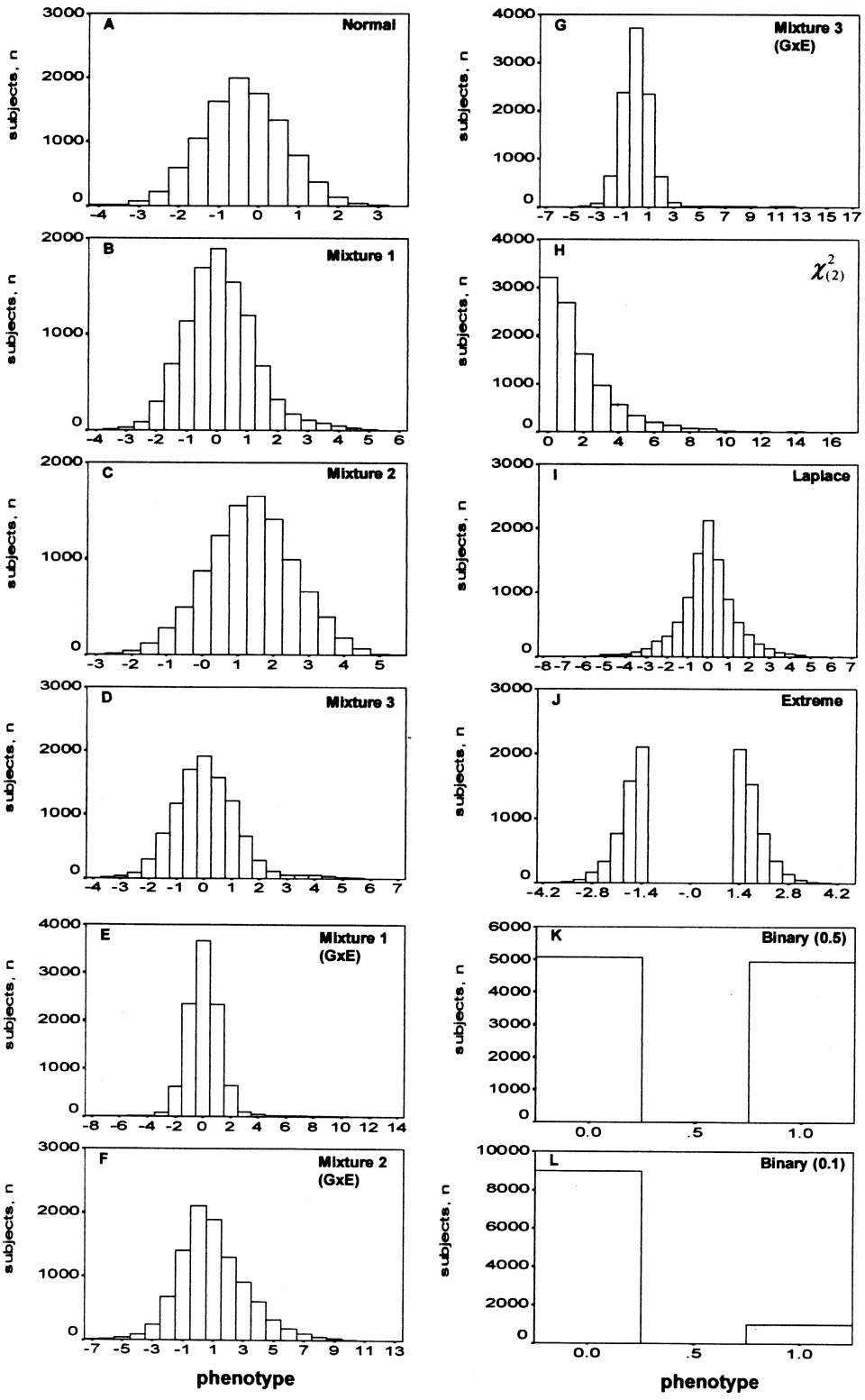

Figure 1 Histograms from a random sample of individuals, for normal distribution (*A*), mixture distributions due to unlinked QTLs (*B–D*), mixture distributions due to G × E (*E–G*), χ^2 distribution with 2 df (*H*), standard Laplace distribution (*I*), extreme sampling from a normal distribution (*J*), and binary distribution with probability of "affection" .5 (*K*) and .1 (*L*). For computational convenience, 10,000 observations were used for each histogram, instead of 100,000.

Table 1

Theoretical Moments for Studied Distributions

Distribution[a]	Mean	Variance	Skewness	Kurtosis
Normal	0	1	0	0
Mixture 1	.118	1.333	.587	1.253
Mixture 2	.000	1.500	.000	−.111
Mixture 3	.088	1.333	.805	2.467
Mixture 1 (G × E)	1	4.667	.755	3.842
Mixture 2 (G × E)	1	5	.537	1.040
Mixture 3 (G × E)	1	4.646	.860	5.346
$\chi^2_{(2)}$	2	4	2	6
Laplace	0	2	0	3
Extreme	0	3.249	0	−1.727
Binary (.5)	.5	.25	0	−2
Binary (.1)	.1	.09	2.67	5.11

[a] For the details of each distribution, see the text.

esis, to test our software and whether the statistical test being used has asymptotically converged to the χ^2 distribution.

Mixture distribution.—To simulate a nonnormal mixture distribution due to a QTL (not at the locus under study), we arbitrarily selected scenarios involving biallelic QTL. In the first experiment (mixture 1), the QTL explained 25% of the phenotypic variance, the increasing allele had a frequency of .20, and its mode of action was recessive. For the second experiment (mixture 2), the QTL explained 33% of the phenotypic variance, both alleles had frequencies of .50, and the mode of action was additive. For the third experiment (mixture 3), the QTL explained 25% of the phenotypic variance, the increasing allele had a frequency of .15, and the mode of action was recessive. In each case, the residual distribution within each genotype class was set to be normal, with a variance of 1.0. Although these effect sizes may be larger than those which typically exist for quantitative traits, they are not larger than those commonly estimated by segregation analyses (Schork et al. 1996), and they allow the effects of nonnormality to be evaluated more clearly.

G × E interaction.—There are many possible models of G × E interaction. To simulate a nonnormal distribution due to G × E interaction, the same mixture models that have been discussed in the immediately preceding subsection were employed, except that the phenotype was computed as $Y = G + R + G * R$ rather than as just $Y = G + R$, where G is the mean value for the genotype at the QTL and R is the residual term. In these cases, the "residual correlation" is expressed as the correlation in the R component, which, for these cases, will not necessarily equal the phenotype correlation after adjustment for the QTL genotype effects. Specifically, when this scheme was applied, G × E mixture 1 had a QTL effect of 7.14% of the variance; G × E mixture 2, 10%; and G × E mixture 3, 7.14%. This had the desirable feature of allowing us to examine the effects of relatively small QTL effects.

"Intrinsic" nonnormality.—To simulate intrinsic nonnormality, we used two models. In the first model, the marginal distribution for each sibling was χ^2 with 2 df; this simulated a markedly skewed distribution. In the second model, the marginal distribution for each sibling was a standard Laplace distribution (Evans et al. 1993); this generated a symmetric but heavy-tailed (leptokurtotic) distribution.

Binary distribution.—To simulate a binary (dichotomous) distribution, we began by simulating an underlying genetic model for a quantitative trait, using the same parameters that were used for mixture 2 described above. We then dichotomized the distribution at a given cut point, to produce a Bernoulli distribution with parameter p, where p is the probability of being affected. Two different distributions were chosen. In the first distribution, p was set equal to .5, to simulate a symmetric distribution; in the second distribution, p was set equal to .1, to simulate a skewed distribution. When the data are dichotomized with p set to .10, for population residual correlations of .10, .30, and .50 at the continuous level, the expected values of the sample phenotypic correlation at the dichotomous level are .09, .16, and .25, respectively. When the data are dichotomized with p set to .50, for population residual correlations of .10, .30, and .50 at the continuous level, the expected values of the sample phenotypic correlation at the dichotomous level are .16, .25, and .34, respectively.

Selective sampling.—To simulate a selective sampling situation, data were selectively sampled from a bivariate normal distribution. Sibling pairs were selected if neither sibling had a trait value between the 10th and 90th percentiles, the so-called EDAC (extreme discordant and concordant) design (Dolan and Boomsma 1998). It is noteworthy that such sampling affects the residual sibling correlation (defined below); specifically, for population correlations of .10, .30, and .50, the expected values of the sample correlation for this extreme sampling are .32, .76, and .92, respectively.

Residual Correlation

By "residual correlation" we mean the correlation between siblings' phenotypes after the effects of genotype at the QTL have been removed. This correlation is a function of both the effects of other genes and the effects of any environmental influence on the trait for which siblings are correlated (Neale and Cardon 1992). Residual correlation affects the power of VC techniques (Schork 1993), and it is possible that it also affects type I error rates when normality is violated. Under the null model of no linkage, the residual correlation equals the phenotypic correlation. Residual correlations of .1, .3,

and .5 were used. We recognize that .5 is a higher correlation than is often observed in sibling data. We include it here to show the effect that strong residual correlation can have. Moreover, although sibling correlations $\geq .5$ are unusual, they have been observed for some common traits, such as height (e.g., see Hopper and Mathews 1983; Keith et al. 1983; Crawford et al. 1993; Christiaens et al. 1997).

Sample Size

Two sample sizes were used—100 sibling pairs and 500 sibling pairs. All sibling pairs were independent; that is, there was only one sibling pair per family.

Generation of Genotypic Data

Because the model used to test for linkage (see the Linkage-Test Statistic subsection, below) uses only IBD status (and not genotype per se), we directly generated an IBD status for each sibling pair, from a binomial distribution with $n = 2$ and $p = \frac{1}{2}$, as would be the case under the null hypothesis. This implies that we are working with perfectly informative markers.

Linkage-Test Statistic

We tested for linkage, using a likelihood-ratio test implemented in a VC model as described by Fulker and Cherny (1996). In brief, this method specifies the values of the expected 2×2 phenotypic-covariance matrix for groups of sibling pairs, as a function of three parameters or VC. Specifically, the expected variances are equal to $\sigma^2_Q + \sigma^2_F + \sigma^2_E$, and the expected covariances are equal to $\pi\sigma^2_Q + \sigma^2_F$, where π is the proportion of alleles that pairs share IBD, σ^2_Q is the variance due to the QTL, σ^2_F represents other sources of variation (both genetic and environmental) that are shared by siblings, and σ^2_E represents sources of variation that are not shared by siblings. The test was also implemented with the raw data (i.e., not summarized as covariance matrices), for the probability that the pair shares zero, one, or two alleles IBD at the marker locus. However, in the current case, this is identical to using the estimated number of alleles IBD, because the markers are perfectly informative. A likelihood-ratio test is conducted to test the significance of σ^2_Q (i.e., under the null hypothesis, no covariation in the trait should be explained by allele sharing, so $H_0{:}\sigma^2_Q = 0$). When data are normally distributed and the variance components are constrained to be nonnegative, the likelihood ratio for this test is distributed, asymptotically under the null hypothesis, as a $\frac{1}{2} : \frac{1}{2}$ mixture of $\chi^2_{(1)}$ and a point mass at zero (Chernoff and Lehmann 1954; Self and Liang 1987).

Linkage-Testing Software

Several software packages are available that allow VC QTL analysis, including Mx (Neale 1997), SOLAR (Almasy and Blangero 1998), and ACT (de Andrade et al. 1997). Prior to conducting our main analysis, we generated five sample data sets of 500 pairs each. We then used these test cases to insure the validity of our software. We independently analyzed each data set by using Mx with the raw data, Mx with the covariance matrices only, ACT, and SOLAR. The likelihood-ratio statistics that we obtained by the four different approaches were very similar, confirming the validity and equivalence of these approaches. After this demonstration, all simulations (except the binary case) were conducted with the Mx software. The binary cases were simulated and evaluated by the SOLAR software. In addition, as a final check, several simulations were done on both Mx and SOLAR.

A FORTRAN program was written to simulate data for the various conditions to be tested. This program used the functions g05ddf and g05caf in the Numerical Algorithms Group library, to obtain pseudorandom numbers from, respectively, the normal and uniform distributions. The simulated data were tested for the presence of a QTL, by an Mx script (Neale 1997). The script, which is given in the Appendix, computed the likelihood of the raw-data vectors, with a model that included effects of a QTL, a residual component shared between siblings, and a residual component not shared between siblings. Twice the difference between the log-likelihood of this model and one in which the QTL effect was omitted was used as a test for the presence of the QTL. This procedure was repeated 100,000 times for each of the models and conditions tested.

Results

Tables 2–7 display the results of the simulation. With normal data (table 2 and figs. 1*A* and 2*A*), which serve as a simple validity check on our simulation software, the type I error rates were consistent with the nominal α levels, for samples of 500. However, for the sample size of 100 and low residual correlation, the error probability was slightly conservative (.03, instead of .05, type I error rates). This result appears to be due to the fact that the sample size is not large enough for the likelihood-ratio test to behave as a χ^2.

Table 3 and figures 1*B–D* and 2*B–D* contain the results for the mixture distributions. As can be seen in table 3, the empirical type I error rates were consistent with the nominal α levels for mixture 2, which was symmetrical and less kurtotic than the normal distribution. However, for mixtures 1 and 3, which were both skewed and leptokurtotic, the type I error rates slightly exceeded

Table 2

Type I Error Rates under Normal Distribution

N and Nominal α Level[a]	Type I Error Rate for Normal Distribution[b]		
	$\rho = .10$	$\rho = .30$	$\rho = .50$
100:			
.05	.02927	.05439	.05585
.01	.00038	.01040	.01188
.001	.00024	.00095	.00127
.0001	.00001	.00012	.00016
500:			
.05	.04407	.05245	.05018
.01	.00736	.01074	.01028
.001	.00038	.00089	.01120
.0001	.00002	.00011	.00009

[a] N = sample size (i.e., no. of sibling pairs) used for each simulated data set.

[b] 100,000 data sets were generated for each parameter set. ρ = residual correlation.

the nominal α levels. Again, the degree of excess was directly related to the residual correlations. With a sibling correlation of .50, the type I error rate was $\leq$.10 for the nominal $\alpha = .05$ level, with the highly skewed and kurtotic mixture 3. The excess error probability was higher for the smaller sample size than for the larger sample size.

Table 4 and figures 1*E* and *F* and 2*E* and *F* contain the results for the mixture distributions due to G × E interaction. As can be seen in table 4, the empirical type I error rates consistently exceeded the nominal α levels for some cases. The degree of excess was directly related to the residual correlations. With sibling correlations of .50, the type I error rates are often double the nominal α level.

With the skewed ($\chi^2_{(2)}$) distribution (table 5 and figs. 1*H* and 2*H*), results consistently show that the empirical type I error rates exceed the nominal α levels by a degree that is directly related to the degree of correlation among the sibling phenotypes. With correlations of .30 and .50, the deviations are substantial. Results for the symmetric but kurtotic Laplace distribution are given in table 5 and figures 1*I* and 2*I* and are extremely similar, in both pattern and magnitude, to those for the χ^2 distribution. Also note that the tails of the distribution of the likelihood-ratio test greatly exceed that predicted by the χ^2 distribution.

With simulation involving extreme sampling and no correction for the extremity (table 6 and figs. 1*J* and 2*J*), results show that the VC statistical procedure employed is quite sensitive to this type of nonnormality. If naively applied without correction to data from extreme samples, it will give far too many false positives when the residual correlation is high. At the .0001 significance level, the type I error rate was $\geq$100-fold higher than that predicted by the χ^2 distribution, for residual correlation $\geq$.30. When the residual correlation is high with the binary distributions (for correlations .5 and .1; table 7 and figs. 1*K* and 2*K*), results show that there is an excess of type I errors that is, again, directly related to the degree of residual correlation.

It is interesting to note the relationship between sample size and significance of the tests (size). For the mixture distributions shown in tables 3 and 4, increasing the sample size considerably decreased the size of the test, specifically for the more extreme significance levels. In contrast, the highly kurtotic distributions show in tables 5 and 6 showed virtually no improvement with increasing sample size.

These results suggest that convergence of the likelihood-ratio test to a limiting χ^2 distribution is, at best, very slow when highly platy- or leptokurtotic distributions are considered. More generally, when the data are not normally distributed, likelihood theory does not insure that the likelihood-ratio test approaches a χ^2 distribution as the sample size increases.

In general, the pattern of results was not dramatically different for different nominal α levels or different sample sizes; the main factors influencing the degree to which the empirical type I error departed from the nominal rate were the marginal phenotypic distribution (specifically the magnitude of leptokurtosis) and residual correlation. This pattern is summarized graphically in figure 2, for 500 sibling pairs and $\alpha = .05$.

One can note that the typical genetically related influences, such as mixtures of normals, do not greatly inflate the type I error unless those mixtures are very extreme and heteroscedastic. By contrast, incorrectly assuming normality when there is marked kurtosis or skewness leads to a significant increase in the type I error rate, in the presence of sibling correlation.

Discussion

Investigators contemplating the use of a likelihood-ratio test in the VC framework, for detection of QTL, can take both comfort and cautions from our results. On the positive side, our validity check shows that three major software packages (Mx, SOLAR, and ACT) yield equivalent answers when applied to the same data. Second, for many of the distributions likely to be encountered by applied investigators, the likelihood-ratio test was remarkably robust. However, on the cautionary side, in some situations the likelihood-ratio test was quite sensitive and resulted in a substantial excess of type I errors.

The current findings show that the likelihood-ratio test in the VC approach to QTL detection considered is robust to some but not all types of deviations from normality that are produced by several different mecha-

Table 3

Type I Error Rates under Mixture Distributions Due to Unlinked QTLs

N and Nominal α Level	Type I Error Rate under								
	Mixture 1[a]			Mixture 2[b]			Mixture 3[c]		
	$\rho = .10$	$\rho = .30$	$\rho = .50$	$\rho = .10$	$\rho = .30$	$\rho = .50$	$\rho = .10$	$\rho = .30$	$\rho = .50$
100:									
.05	.05819	.06775	.07797	.02900	.05059	.05634	.04457	.07038	.09777
.01	.01213	.01681	.02209	.00038	.00910	.01170	.00913	.01855	.03428
.001	.00137	.00211	.00413	.00025	.00840	.00120	.00092	.00282	.00827
.0001	.00011	.00032	.00075	.00001	.00005	.00009	.00010	.00031	.00207
500:									
.05	.06321	.06449	.06851	.04438	.05295	.04904	.06169	.07085	.08570
.01	.01466	.01600	.01777	.00707	.01113	.00982	.01360	.01847	.02756
.001	.00192	.00221	.00276	.00056	.00102	.00110	.00148	.00318	.00571
.0001	.00018	.00034	.00048	.00003	.00008	.00004	.00012	.00051	.00145

Note.—See footnotes to table 2.

[a] "Increasing" allele frequency = .20; QTL explains 25% of the phenotypic variance; mode of action is recessive.

[b] "Increasing" allele frequency = .50; QTL explains 33% of the phenotypic variance; mode of action is additive.

[c] "Increasing" allele frequency = .15; QTL explains 25% of the phenotypic variance; mode of action is recessive; 100,000 data sets were generated for each parameter set.

nisms. The degree of type I error–rate inflation was directly related to the residual phenotypic sibling correlation. With correlations $\leq .50$, the type I error rates at the $\alpha = .05$ were $\leq .20$ for some models. When excessive type I error rates did occur, for samples sizes ≤ 500 sibling pairs there was no great diminution of this effect for markedly nonnormal distributions. These results are quite consistent with findings from the general structural equation–modeling literature, showing a 1.5–2-fold inflation of the type I error rate under modestly kurtotic distributions, with sample sizes <1,000 (Hu and Bentler 1992). On the other hand, for more modest forms of nonnormality and lower residual correlation, the likelihood-ratio test with the VC method appeared to be quite robust.

It is possible that investigators are used to assuming that the statistical tests that they employ are highly robust to violations of normality, particularly when the sample size includes ≥ 100 observations. However, this confidence may come from robustness studies involving tests of means. Nearly 50 years ago, Box (1953, p. 318) recognized that tests on variances were less robust. He stated that "it would be appear, however, that this remarkable property of 'robustness' to non-normality which these tests for comparing means possess, and without which they would be much less appropriate to the needs of the experimenter, is not necessarily shared by other statistical tests, and in particular is not shared by tests for equality of variances."

There is a large body of statistical theory concerning both the validity of the likelihood-ratio test under model misspecification and available options for making inferences more robust (e.g., see Foutz and Srivastava 1977; White 1994). In the context of VC tests, it is clear that nonzero kurtosis is a primary culprit leading to deviations of the test statistic from its asymptotic distribution (Beaty et al. 1985). A number of alternative tests based on simple corrections of either the likelihood-ratio test or score tests are available (Beaty et al. 1985; White 1994) and could serve as alternatives to the classic likelihood-ratio test when multivariate normality is grossly violated.

It is important to consider the conditions of these simulation studies, to place them in context. These simulations involve sibling pairs, perfectly informative markers, $\alpha = .05$–$.0001$, sample sizes of 100–500, sibling correlations of .10–.50, and particular types of nonnormality. The extent to which our results apply to other sampling units (e.g., large sibships or pedigrees), other types of VC-based tests (Amos et al. 1996), less-informative markers, or different nominal α levels, sample sizes, sibling correlations, and types of nonnormality remains open to question. It is especially interesting to determine whether results would generalize to the more common case of partially informative markers. Whether such cases would ameliorate or exacerbate the effects of marked nonnormality remains unknown.

Some of the types of nonnormality that we have studied may be more extreme than those commonly encountered in practice. However, others may be quite typical. For example, in two large population-based cohorts that we have reported elsewhere (Allison et al., in press), body-mass index (BMI [kg/m^2]) had skewness coefficients of 1.0–1.2 and kurtosis coefficients of 2.5–3.3. Similarly, the third National Health and Nutrition Examination Survey (National Center for Health Statistics 1997), a nationally representative sample of the U.S. population in whom weight and height were meticu-

Table 4

Type I Error Rates under Mixture Distributions Due to G × E

N and Nominal α Level	Type I Error Rates under								
	Mixture 1 (G × E)[a]			Mixture 2 (G × E)[b]			Mixture 3 (G × E)[c]		
	$\rho = .10$	$\rho = .30$	$\rho = .50$	$\rho = .10$	$\rho = .30$	$\rho = .50$	$\rho = .10$	$\rho = .30$	$\rho = .50$
100:									
.05	.06757	.10015	.12626	.03152	.06053	.08561	.06939	.12422	.19220
.01	.02059	.03832	.05588	.00586	.01473	.02613	.03042	.06146	.11232
.001	.00455	.01053	.01844	.00077	.00232	.00490	.01015	.02408	.05485
.0001	.00111	.00295	.00639	.00007	.00037	.00083	.00347	.00944	.02588
500:									
.05	.08431	.10145	.11297	.04543	.06472	.07915	.08551	.12002	.15783
.01	.02528	.03532	.04393	.00784	.01653	.02301	.02885	.05367	.08278
.001	.00433	.00809	.01207	.00071	.00207	.00429	.00778	.01830	.03547
.0001	.00079	.00191	.00344	.00007	.00028	.00007	.00254	.00665	.01653

Note.—See footnotes to table 2.

[a] "Increasing" allele frequency = .20; mode of action is recessive.

[b] "Increasing" allele frequency = .50; mode of action is additive.

[c] "Increasing" allele frequency = .15; mode of action is recessive; 100,000 data sets were generated for each parameter set.

lously measured, had, for BMI, skewness and kurtosis coefficients that were 1.2 and 3.4, respectively, even after adjustment for age and sex (National Center for Health Statistics 1997). These values are quite similar to those for some of the models that we have simulated.

The combination of nonnormality and positive sibling correlation (although perhaps not as high as .50) may be common for several reasons. First, apart from the seemingly unlikely circumstance that siblings are negatively correlated for environmental influences on the phenotype, the presence of some genetic influence on the trait virtually assures that there will be some sibling correlation. If siblings were not correlated for the trait, it is unlikely that a QTL-mapping study would ever be initiated. Second, if there are any "major" genes or "oligogenes" for the trait under study, then this would also produce nonnormality (Lander and Botstein 1989; Wright and Kong 1997), and, again, one typically enters a QTL-mapping study only when there is a strong priori reason to suspect major genes or oligogenes. Finally, independent of any particular theorizing about genetic effects, it has been noted that, for whatever reason, many human traits are not normally distributed (Micceri 1989). Thus, the issue raised herein may apply to many QTL-mapping situations.

In this regard, the nonnormality of both the Laplace distribution and other kurtotic but symmetric distributions is of special interest. It points out that analysts should not rely on the apparent symmetry of a distribution as being indicative of normality sufficient for ap-

Table 5

Type I Error Rates under $\chi^2_{(2)}$ and Laplace Distributions

N and Nominal α Level	Type I Error Rate under					
	$\chi^2_{(2)}$			Laplace		
	$\rho = .10$	$\rho = .30$	$\rho = .50$	$\rho = .10$	$\rho = .30$	$\rho = .50$
100:						
.05	.04933	.11125	.18658	.04521	.11550	.17290
.01	.01222	.04117	.09986	.01036	.04423	.08901
.001	.00228	.01003	.04037	.00131	.01154	.03562
.0001	.00037	.00270	.01584	.00002	.00292	.01416
500:						
.05	.06824	.11860	.18192	.06674	.11800	.17031
.01	.01774	.04681	.09957	.01590	.04784	.08789
.001	.00258	.01266	.04366	.00215	.01391	.03564
.0001	.00039	.00346	.02040	.00044	.00431	.01464

Note.—See footnotes to table 2.

Table 6

Type I Error Rates under Extreme Sampling Distribution

N and Nominal α Level	Type I Error Rate under Extreme Sampling Distribution		
	$\rho = .10$	$\rho = .30$	$\rho = .50$
100:			
.05	.07316	.17313	.20787
.01	.02101	.09528	.11618
.001	.00397	.04541	.04820
.0001	.00108	.02382	.01766
500:			
.05	.06441	.15371	.19736
.01	.01643	.07488	.11401
.001	.00229	.02891	.05388
.0001	.00028	.01123	.02575

NOTE.—See footnotes to table 2.

plication of asymptotic inference (Bentler et al. 1991). Clearly, the nonnormality of the symmetric Laplace-distribution example led to gross violations in the presence of high residual correlations. The reason for this appears to be the combination of (*a*) overweighting of trait values in the tail, beyond that expected by normality, and (*b*) a sibling correlation that makes it likely that both values in the sibship will occur in the same area of distribution. Under such circumstances, random IBD configurations that by chance co-occur with higher sharing for such pairs will provide LOD scores that are inflated. Such distortions might be detected in the context of a genome scan, if they lead to an unusually large number of large-LOD-score peaks.

Simple transformation will often be sufficient to normalize a nonnormal distribution to such an extent that inference is valid. However, such transformation cannot be assumed to always work. For example, if the χ^2 distribution depicted in figure 1*H* were transformed via a log transformation, the data would still be significantly skewed and kurtotic. Of course, more-general transformations are available (Box and Cox 1964; George and Elston 1988), but even these are not guaranteed to induce normality. If the distribution is continuous, an inverse Gaussian transformation based on ranks can always be relied on for normalization, although the new reliance on order statistics may create additional issues and potentially decrease power (Wilcox 1997).

The point that the presence of a major gene or oligogene generates a nonnormal mixture distribution has made been before (Lander and Botstein 1989; Schork et al. 1990*a*, 1990*b*, 1996; Amos et al. 1996) but seems not to be widely appreciated. The fundamental quandary in the use of likelihood ratio–based tests of allele sharing for linkage analyses in VC models is that one is typically relying on second-order statistics—that is, variances and covariances—to model a phenomenon (i.e., the presence of a linked gene) that also influences first-order statistics (i.e., means) of the trait distribution. Consider the fact that, if a locus influences a quantitative trait, then individuals with different genotypes will have trait values that cluster around different mean values, thus creating a "mixture" of individuals in the population (i.e., those with different genotypes). The mean effects and frequency of these genotypes will then create (potentially) markedly nonnormal trait distributions, although the distribution of trait values *within* a particular genotypic category could be normal. By relying on test statistics that do not consider the effects that underlying genotypes have on mean trait values but that, with respect to genotype and phenotype, use only variances or covariances among relatives, one may require, for relevant test-statistic formulation, normality of the overall trait distribution in the population at large. This requirement stands in contrast to the potential existence of a locus

Table 7

Type I Error Rates under Binary Distributions

N and Nominal α Level	Type I Error Rate under					
	Binary (.5)			Binary (.1)		
	$\rho = .10$	$\rho = .30$	$\rho = .50$	$\rho = .10$	$\rho = .30$	$\rho = .50$
100:						
.05	.04728	.06396	.08004	.07785	.07405	.10100
.01	.01015	.01728	.02643	.02775	.04581	.06663
.001	.00150	.00339	.00617	.00823	.01551	.02339
.0001	.00027	.00078	.00204	.00244	.00486	.00904
500:						
.05	.05587	.05992	.04094	.08906	.12165	.15087
.01	.01362	.01356	.01935	.03068	.05069	.07403
.001	.00162	.00208	.00368	.00937	.01528	.02672
.0001	.00011	.00029	.00068	.00433	.00534	.01190

NOTE.—See footnotes to table 2.

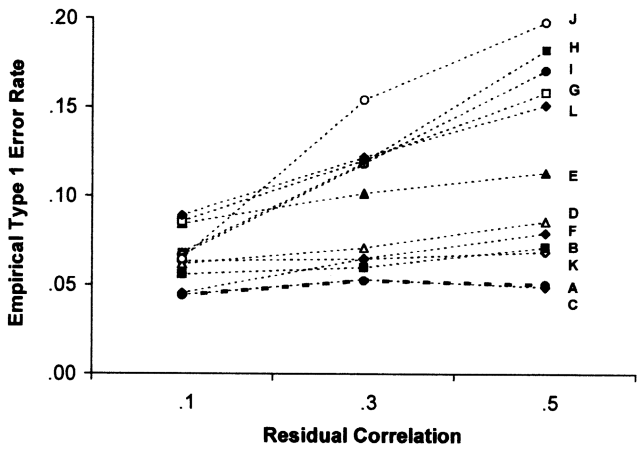

Figure 2 Empirical type I error rate from 100,000 samples of 500 sib pairs, at $\alpha = .05$, for normal distribution (*A*), mixture distributions due to unlinked QTLs (*B–D*), mixture distributions due to G × E (*E–G*) χ^2 distribution with 2 df (*H*), standard Laplace distribution (*I*), extreme sampling from a normal distribution (*J*), and binary distribution with probability of "affection" .5 (*K*) and .1 (*L*).

effect and can create problems. This point requires greater recognition. Linkage tests that simultaneously accommodate the VC but also account for a mixture distribution by modeling the mean genotype effects may be useful in addressing this problem (Schork 1992), but that approach may be computationally prohibitive (see Schork 1991). In our study, the presence of a mixture of normals (mixtures 1–3) did not greatly increase the type I error rate until they became extreme and were combined with heteroscedasticity.

With regard to VC, only when it can be established that the asymptotic distribution of this test statistic holds for the application being considered does it seem to be prudent to make inferences based on the likelihood-ratio test. Fortunately, it is possible to make such an assessment via simulation, and methods exist for doing so (e.g., see Iturria et al., in press). If it is demonstrated that the asymptotic test distribution does not hold, then alternative approaches, involving modifications of the VC method, to make it more robust, can be employed. These approaches may include alternative estimation methods, such as generalized estimation equations (Amos 1994) or least squares (Goldstein 1994; Elston et al., in press); alternative robust test statistics, such as the robust-score test (Beaty et al. 1985); utilization of a multivariate *t* distribution that allows fatter tails (Lange et al. 1989); or computationally intensive permutation-based testing (Guerra et al., in press; Iturria et al., in press).

A second potential response to marked nonnormality would be to return to more-robust procedures utilizing only pairwise differences between siblings, such as the traditional Haseman-Elston test (Haseman and Elston 1972; Wan et al. 1997) or the nonparametric analogue of it (Kruglyak and Lander 1995). Given the relatively lower power of pair-difference–based methods compared with the VC approach (Wright 1997; Williams and Blangero 1999), such a choice may be suboptimal.

A third option is to transform the data to approximate normality, as has been discussed above. In many cases this will work quite well, especially either when the source of nonnormality is scale dependent or in some of the G × E models that we have simulated. However, there is no guarantee that a transformation exists that will make a particular error distribution normal. Moreover, it is possible that, under the alternative hypothesis, such a procedure might decrease power or obscure other trait-relevant phenomena. As stated above (see the Introduction section), the presence of a QTL induces a mixture distribution, and forcing the data to be normally distributed might remove some of the apparent effect of the QTL (Schork and Schork 1989; Schork et al. 1996).

A fourth option would be to abandon the assumption that the test statistic (twice the natural log of the likelihood ratio) has a particular distribution (i.e., a $\frac{1}{2} : \frac{1}{2}$ mixture of $\chi^2_{(1)}$ and 0) and, instead, derive critical values of the test statistic via distribution-free resampling–based techniques (Schork and Schork 1989; Dunn et al. 1993; Good 1994; Guerra et al., in press), as discussed above. Although theoretically sound, this approach could be computationally demanding.

Fifth, in some cases considered that we have considered, it is possible to alter the likelihood calculation, to account for the nonnormality. This is the case for nonrandom sampling, and, within Mx, SOLAR and ACT, weights or ascertainment correction could be used, provided that the sampling scheme is known. Alternatively, data from nongenotyped sibling pairs used to screen the selected pairs could be included in the analysis, if we assume that the data are missing at random (Little and Rubin 1987; Neale, in press). For the G × E cases, it may also be possible to model the G × E effect explicitly and to reduce the problem to distributions that are normal when conditioned on the genotypic or environmental values.

Finally, some investigators (Astemborski et al. 1985; Beaty et al. 1985, 1987*a,* 1987*b;* Amos et al. 1996) have proposed "robust" VC methods, based on quasi-likelihood, that should be less dependent on the normality assumption as discussed above.

In conclusion, results that have been provided herein clearly show that the likelihood-ratio test within the VC QTL-detection procedure studied is robust to some types of nonnormality but is not so to others. Our results certainly indicate that blindly applying the likelihood-ratio test in a VC QTL-mapping analysis, without regard

to the phenotypic distribution, is an unsound practice. However, we wish to emphasize that we believe that VC models are among our most powerful tools. Thus, we are suggesting both a careful examination of distributions prior to analysis and, when needed, application of modifications/corrections to the standard VC implementation, as has been discussed above (e.g., see Amos et al. 1996; Guerra et al., in press; Wang et al., in press). Several alternatives that preserve the power of the VC approach under the alternative hypothesis but offer better control of the type I error rate under nonnormality have been discussed. Further research is needed to evaluate, empirically, the performance of these techniques.

Acknowledgments

This research was supported in part by National Institutes of Health grants DK47256, DK51716, DK26687, MH01458, GM18897, HL45522, MH59490, GM52607, and ES09912 and by a grant from Gemini Corporation.

Appendix

Mx Script

```
! Mx script to test equality of covariances in 3 groups
! Any characters after a "!" on a line are comments
! Three groups
#Ngroups 3
Group 1: IBD 2 cases
Data NInput=3 ! Three variables to be analyzed
Labels IBD Sib1 Sib2 ! Names for variables
Rectangular file=simnor.rec ! Read data file
! The file simnor.rec contains ibd sib1 sib2 records of the form
! 1 −4.770278154788629E−002 −0.818980724789924
! 2 −6.528038280488491E−002 −5.055716933976841E−002
! 2 −2.196692932777173E−002 −0.426106954533204
! 1 −1.24327779081014 −0.212518588249240
! 0 −0.604940258106770 2.31134951141858
Select if ibd = 2 ! Select IBD=2 cases
Select Sib1 Sib2; ! Drop ibd from analysis
Begin Matrices;
X Full 1 1 Free ! Additive QTL variance component
Y Full 1 1 Free ! Residual shared covariance component
Z Full 1 1 Free ! Residual nonshared variance component
M full 1 2 Free ! Free parameters for means
End Matrices;
Specify M 4 4 ! Equate mean parameters for sib1 and sib2
Start 1 all ! Start all free parameters at 1.0
Begin Algebra;
Q=X*X'; ! Compute squared QTL component
R=Y*Y'; ! Compute squared residual shared component
E=Z*Z'; ! Compute squared residual nonshared component
End Algebra;
Covariance Q+R+E | Q+R_! Construct 2 × 2 predicted covariance matrix
Q+R | Q+R+E; ! To fit to sib1,sib2 phenotypes
Means M; ! Predicted means
End Group;
Group 2: IBD 1 cases ! Second group, commands as above
Data NInput=3
Labels IBD Sib1 Sib2
Rectangular file=simnor.rec
Select if IBD=1
Select Sib1 Sib2;
```

```
Begin Matrices=Group 1; ! Equate all matrices to those of group 1
H Full 1 1 ! .5 constant
End Matrices;
Matrix H .5
Covariance Q+R+E | h@Q+R_! IBD 1 share half of QTL covariance
h@Q+R | Q+R+E;
Means M;
End Group;
Group 3: IBD 0 pairs
Data NInput=3
Labels IBD Sib1 Sib2
Rectangular file=simnor.rec
Select if ibd=0
Select Sib1 Sib2;
Begin Matrices=Group 1;
End Matrices;
Means M;
Covariance Q+R+E | R_! No QTL covariance for IBD 0
R | Q+R+E;
Option nd=6 ! 6 decimal places of precision
Option Issat ! This is full, or a saturated model
Option Multiple ! Prepare to fit submodel next
End Group ! End of first job
Drop X 1 1 1 ! Drop (fix at 0) QTL component
End ! End of submodel job
```

Electronic-Database Information

The URL for data in this article is as follows:

Numerical Algorithms Group, http://www.nag.co.uk (for functions g05ddf and g05caf)

References

Allison DB, Heo M, Schork NJ, Wong SL, Elston RC (1998) Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power. Hum Hered 48:97–107

Allison DB, Schork NJ (1997) Selected methodological issues in meiotic mapping of obesity genes in humans: issues of power and efficiency. Behav Genet 27:401–421

Allison DB, Zannolli R, Faith MS, Heo M, Pietrobelli A, Vanltallie TB, Pi-Sunyer FX, et al (1999) Weight loss increases and fat loss decreases all-cause mortality rate: results from two independent cohort studies. Int J Obesity 23:603–611

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62: 1198–1211

Amos CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 54: 535–543

Amos CI, Krushkal J, Thiel TJ, Young A, Zhu DK, Boerwinkle E, de Andrade M (1997) Comparison of model-free linkage mapping strategies for the study of a complex trait. Genet Epidemiol 14:743–748

Amos CI, Zhu DK, Boerwinkle E (1996) Assessing genetic linkage and association with robust components of variance approaches. Ann Hum Genet 60:143–160

Astemborski JA, Beaty TH, Cohen BH (1985) Variance components analysis on forced expiration in families. Am J Med Genet 21:741–753

Beaty TH, Liang KY (1987*a*) Robust inference for variance components models in families ascertained through probands. I. Conditioning on proband's phenotype. Genet Epidemiol 4:203–210

Beaty TH, Liang KY, Seerey S, Cohen BH (1987*b*) Robust inference for variance. Genet Epidemiol 4:211–221

Beaty Th, Self SG, Liang KY, Connolly MA, Chase GA, Kwiterovich PO (1985) Use of robust variance components models to analyses triglyceride data in families. Ann Hum Genet 49:315–328

Bentler PM, Maia B, Yukata K (1991) Covariance structure analysis under a simple kurtosis model. In: Keramidas EM (ed) Computing science and statistics. Proceeding of the 23rd Symposium on the Interface. Interface Foundation of North America, Fairfax Station, VA, pp 463–465

Box GEP (1953) Non-normality and tests on variances. Biometrika 40:318–335

Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc 26:211–252

Chernoff H, Lehmann EL (1954) The use of maximum likelihood estimates in χ^2 tests for goodness of fit. Ann Math Stat 25:579–586

Christiaens GC, Nieuwenhuis HK, Bussel JB (1997) Comparison of platelet counts in first and second newborns of mothers with immune thrombocytopenic purpura. Obstet Gynecol 90:546–552

Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, Stern MP, et al (1997) A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. Nat Genet 15:273–276

Crawford DH, Halliday JW, Summers KM, Bourke MJ, Powell LW (1993) Concordance of iron storage in siblings with genetic hemochromatosis: evidence for a predominantly genetic effect on iron storage. Hepatology 17:833–837

de Andrade M, Thiel TJ, Yu LP, Amos CI (1997) Assessing linkage on chromosome 5 using components of variance approach: univariate versus multivariate. Genet Epidemiol 14:773–778

Dolan CV, Boomsma DI (1998) Optimal selection of sib pairs from random samples for linkage analysis of a QTL using the EDAC Test. Behav Genet 28:197–206

Duggirala R, Williams JT, Williams-Blangero S, Blangero J (1997) A variance component approach to dichotomous trait linkage analysis using a threshold model. Genet Epidemiol 14:987–992

Dunn G, Everitt B, Pickles A (eds) (1993) Modeling covariances and latent variables using EQS. Chapman & Hall, London

Eaves L, Meyer J (1994) Locating human quantitative trait loci: guidelines for the selection of sibling pairs for genotyping. Behav Genet 24:443–455

Eaves LJ, Neale MC, Maes H (1996) Multivariate multipoint linkage analysis of quantitative trait loci. Behav Genet 26: 519–525

Elston RC, Buxbaum S, Jacobs KB, Olson JM. Haseman and Elston revisited. Genet Epidemiol (in press)

Evans M, Hatings N, Peacock B (1993) Laplace distribution. In: Evans M, Hatings N, Peacock B (eds) Statistical distributions. Wiley-Interscience, John Wiley & Sons, New York, pp 92–94

Foutz RV, Srivastava RC (1977) The performance of the likelihood test when the model is incorrect. Ann Stat 5: 1183–1194

Fulker DW, Cherny SS (1996) An improved multipoint sib-pair analysis of quantitative traits. Behav Genet 26:527–532

George VT, Elston RC (1988) Generalized modulus power transformations. Commun Stat Theory Methods 17: 2933–2952

Goldstein H (1994) Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika 73: 43–56

Good P (1994) Permutation tests: a practical guide to resampling methods for testing hypothesis. Springer-Verlag, New York

Guerra R, Wan Y, Jia A, Amos CI, Cohen JC (1999) Testing for linkage under robust genetic models. Hum Hered 49: 146–153

Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 2:3–19

Hopper JL, Mathews JD (1983) Extensions to multivariate normal models for pedigree analysis. II. Modeling the effect of shared environment in the analysis of variation in blood lead levels. Am J Epidemiol 117:344–355

Hu L, Bentler PM (1992) Can test statistics in covariance structure analysis be trusted? Psychol Bull 112:351–362

Iturria SJ, Dyer TD, Williams JT, Blangero J. An empirical test of the significance of an observed QTL effect that preserves additive genetic variance. Genet Epidemiol (in press)

Keith RA, Van Loon J, Wussow LF, Weinshilboum RM (1983) Thiol methylation pharmacogenetics: heritability of human erythrocyte thiol methyltransferase activity. Clin Pharmacol Ther 34:521–528

Kruglyak L, Lander ES (1995) A nonparametric approach for mapping quantitative trait loci. Genetics 139:1421–1428

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Lange KL, Little RJA, Taylor JMG (1989) Robust statistical modeling using the t distribution. J Am Stat Assoc 85: 881–896

Little RJA, Rubin DB (1987) Statistical analysis with missing data. John Wiley & Sons, New York

Matthysse S, Lange K, Wagener DK (1979) Continuous variation caused by genes with graduated effects. Proc Nat Acad Sci USA 76:2862–2865

Micceri T (1989) The unicorn, the normal curve, and other improbable creatures. Psychol Bull 105:156–166

Motum PI, Donald JA, Trent RJ (1993) Linkage analysis of the hemoglobin F determinant(s) in an Australian hemoglobin Lepore (Boston) kindred. Am J Hematol 43:37–43

National Center for Health Statistics (1997) National Health and Nutrition Examination Survey III, 1988–1994. National Center for Health Statistics, Centers for Disease Control and Prevention, Rockville, MD

Neale MC (1997) Mx: statistical modeling, 4th ed. Department of Psychiatry, Medical College of Virginia, Richmond

——— Individual fit, heterogeneity, and missing data in longitudinal and multilevel structural equation modeling. In: Little T, Schnabel K (eds) Lawrence Erlbaum, New York (in press)

Neale MC, Cardon RL (1992) Methodology for genetic studies of twins and families. In: Neale MC, Cardon LR (eds) NATO ASI, series D. Vol 47: Behavioral and social sciences. Kluwer Academic, Dordrecht, Boston, London

Pooni HS, Jinks JL (1976) The efficiency and optimal size of triple test cross design for detecting epistatic variation. Heredity 36:215–227

Pratley RE, Thompson DB, Prochazka M, Baier L, Mott D, Ravussin E, Sakul H, et al (1998) An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians. J Clin Invest 101:1757–1764

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Risch N, Zhang H (1995) Extreme discordant sib-pairs for mapping quantitative trait loci in humans. Science 268: 1584–1589

——— (1996) Mapping quantitative trait loci with extreme discordant sib pairs: sampling considerations. Am J Hum Genet 58:836–843

Schork NJ (1991) Efficient computation of patterned covariance matrix mixed model in quantitative segregation analysis. Genet Epidemiol 8:29–46

——— (1992) Extended pedigree patterned covariance matrix mixed models for quantitative phenotype analysis. Genet Epidemiol 9:73–86

——— (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. Am J Hum Genet 53:1306–1319

Schork NJ, Allison DB, Thiel B (1996) Mixture distributions in human genetics research. Stat Methods Med Res 5: 155–178

Schork N, Schork MA (1989) Testing separate families of segregation hypotheses: bootstrap methods. Am J Hum Genet 45:803–813

Schork NJ, Weder AB, Schork MA (1990*a*) On the asymmetry of biological frequency distributions. Genet Epidemiol 7: 427–446

Schork NJ, Weder AB, Schork MA, Bassett DR, Julius S (1990*b*) Disease entities, mixed multi-normal distributions, and the role of the hyperkinetic state in the pathogenesis of hypertension. Stat Med 9:301–314

Self SG, Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Wan Y, Cohen J, Guerra R (1997) A permutation test for the robust sib-pair linkage method. Ann Hum Genet 61:79–87

Wang J, Guerra R, Cohen J (1998) Statistically robust approaches for sib-pair linkage analysis. Ann Hum Genet 62: 349–359

White H (994) Estimation, inference and specification analysis. Cambridge University Press, Cambridge

Wijsman EM, Amos CI (1997) Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. Genet Epidemiol 14: 719–735

Wilcox RR (1997) Introduction to robust estimation and hypothesis testing. Academic Press, San Diego

Williams JT, Blangero J (1999) Comparison of variance components and sibpair-based approaches to quantitative trait linkage analysis in unselected samples. Genet Epidemiol 16: 113–134

Wright FA (1997) The phenotypic difference discards sib-pair QTL linkage information. Am J Hum Genet 60:740–742

Wright FA, Kong A (1997) Linkage mapping in experimental crosses: the robustness of single gene models. Genetics 146: 417–425