

# Detection of Genetic Heterogeneity for Complex Quantitative Phenotypes

Nicholas J. Schork

*Department of Medicine and Department of Epidemiology, University of Michigan, Ann Arbor, Michigan*

Statistically characterizing factors responsible for quantitative phenotype expression (e.g., polygenes, major genes, shared household factors, etc.) through model selection strategies is a difficult task. A great deal of effort has been expended on refining mathematical and computational aspects of various segregation models used to characterize unique expressions of quantitative phenotypes in an effort to make these models easier to implement and evaluate for a given set of data. In this paper a slightly different angle is emphasized: namely, the explicit modeling of the potentially numerous heterogeneous genetic and environmental processes (i.e., segregation patterns, household aggregations, etiologic processes, etc.) that could contribute to the overall variation of a quantitative trait. As such, this paper describes tools for detecting quantitative trait heterogeneity that are meant to answer such questions as, “are there pedigrees among a great many that show a pattern consistent with a possibly very specific single locus segregation pattern while the rest show compatibility with a polygenic or purely environmental pattern?” Methods for determining the significance of such heterogeneity are also discussed, as are the results of numerous examples and simulation studies carried out in an effort to validate and further elaborate aspects of the proposed techniques. © 1992 Wiley-Liss, Inc.

**Key words:** genetic heterogeneity, quantitative phenotypes, segregation analysis, measured genotype, linkage analysis

## INTRODUCTION

The problem of determining the genetic basis of human quantitative phenotype expression through statistical analysis has been a thorn in the side of statistical geneticists for some time. Karl Pearson [1893, 1895] first noted that the distributional forms a quantitative trait can exhibit in the population at large are compatible with a number of

Received for publication December 2, 1991; revision accepted May 7, 1992.

Address reprint requests to Nicholas J. Schork, Department of Medicine and Department of Epidemiology, University of Michigan, R6592 Kresge I, Ann Arbor, MI 48109-0500.

© 1992 Wiley-Liss, Inc.

different and often hard to discern hypotheses one could put forth about the determinants of the trait (see Schork, et al. [1990b]). Many current distributional modeling strategies for quantitative phenotype analysis typically trade a comprehensiveness in scope (e.g., by allowing for such things as the simultaneous estimation of polygenic parameters and parameters associated with a single locus with large effects) for computational practicality, as is evidenced by the notoriously difficult "mixed model" for quantitative traits [Elston, 1981]. Although a great deal of sophistication—both analytical [Ott, 1979; Hasstedt, 1982; Bonney, 1984; Schork, 1991a] and computational [Demenais et al., 1990; Schork, 1991b,c]—has softened the intimidating demands of the mixed model, its primary assumption, that highly specific major locus, polygenic, and possibly shared environmental effects, are ubiquitous with respect to both normal and deleterious manifestations of a trait and are, in fact, recognizable and determinable, in some sense, in everyone, can and should be called into question.

In this paper an alternative to traditional mixed model-like strategies is offered which, although not particularly easy to implement on a computer, does afford a great deal of flexibility and may be more appropriate than the classical mixed model for a wide variety of traits. The proposed techniques, which can be collectively described as "admixed models" following Greenberg and Hodge [1985], and as opposed to the traditional "mixed models," work by assuming that the trait in question is manifested in different *forms* in different *groups of people* for different *reasons*. A goal of the proposed methodology is thus to estimate not only parameters associated with particular characterizations of the different etiologic processes contributing to the variation and expression of the phenotype, but also to estimate the very frequencies of these various processes as well. The proposed methods can be considered extensions of the models for linkage analyses involving qualitative traits put forth by Smith [1963], Ott [1977], Risch and Baron [1982], and Greenberg and Hodge [1985].

Quantitative traits amenable to the methods proposed herein abound, although blood pressure can be seen as more or less paradigmatic of the type of trait which might benefit most from the proposed methodology: a veritable plethora of associations, predictors, markers, secondary causes, etc. for blood pressure and its disease manifestations have been reported that simply defies the development of a single, comprehensive framework from which the variation of blood pressure values can be viewed as a whole. As an alternative to the futile creation of a comprehensive framework, Schork et al. [1990a] and Weder et al. [1989, 1991] advocate both the identification of pedigrees with similar etiologic patterns as well as the identification of the genetic and physiologic causes of the patterns themselves in an effort to foster further study of each individual etiologic or pathophysiologic mechanism.

## MATERIALS AND METHODS

### Admixed Models

A description of the admixture model approach to characterizing heterogeneous quantitative phenotypes is straightforward once some notational details have been worked out. Let  $\theta_i$  denote the parameters associated with a particular genetic model  $i$ , with density function  $f_i$ . Likelihoods,  $L(\theta|X)$ , given  $N$  pedigrees whose data can be collectively represented by  $X = \{x_1, \dots, x_N\}$ , can then be easily constructed; for example, one can write

$$\begin{aligned}
 L_m(\theta_m|X) &= \prod_{i=1}^N f_m(x_i|p_A, \mu_{AA}, \mu_{Aa}, \mu_{aa}, \sigma_c^2) \\
 L_p(\theta_p|X) &= \prod_{i=1}^N f_p(x_i|\mu, \sigma_a^2, \sigma_d^2, \sigma_e^2) \\
 L_h(\theta_h|X) &= \prod_{i=1}^N f_h(x_i|\mu, \sigma_a^2, \sigma_h^2, \sigma_e^2) \\
 L_a(\theta_a|X) &= \prod_{i=1}^N f_a(x_i|\beta, \sigma_h^2, \sigma_e^2)
 \end{aligned} \tag{1}$$

where (subscript)  $m$  denotes a 2-allele major locus model with allele frequency parameter  $p_A$ , mean genotype effect parameters  $\mu_{AA}$ ,  $\mu_{Aa}$ ,  $\mu_{aa}$ , and variance term  $\sigma_c^2$  [Elston et al., 1975; Elston, 1981];  $p$  denotes a polygenic model with mean  $\mu$ , dominance variance  $\sigma_d^2$ , additive variance  $\sigma_a^2$ , and environmental variance  $\sigma_e^2$  [Lange et al., 1974; Schork and Schork, 1991];  $h$  denotes a shared household model with shared household variance  $\sigma_h^2$ , additive genetic variance  $\sigma_a^2$ , and environmental variance  $\sigma_e^2$  [Moll et al., 1984]; and  $a$  is an age-mediated, shared household model with age associated regression coefficient  $\beta$ , shared household variance  $\sigma_h^2$ , and environmental variance  $\sigma_e^2$ . This last model would be compatible with hypotheses claiming that a certain trait evolves as a result of, say, prolonged dietary habits (possibly among family members) or through variations in environmental exposures to certain toxic agents (e.g., cigarette use among family members). For present purposes we will forego detailed elaboration of the derivations and constructions of the models in (1) and instead invite the reader to peruse the relevant references. In addition, it should be emphasized that the models in (1) are far from exhaustive and have been chosen merely to characterize the wide variety of models one may use to represent different etiologic or developmental patterns for quantitative traits.

To model the etiologic and genetic heterogeneity of a trait, we consider estimating parameters consistent with different models *simultaneously* with parameters included to account for the *frequencies* with which the processes underlying the models occur. Thus, if it is assumed that a disease “level” or manifestation of a quantitative trait (e.g., as hypertension is to blood pressure) is affected by a single locus with large effects *or* by a complex of genetic loci (i.e., polygenes), one could write down a likelihood function as

$$L_{m,p}(\alpha, \theta_m, \theta_p|X) = \prod_{i=1}^N [\alpha \cdot f_m(x_i|\theta_m)] + [(1 - \alpha) \cdot f_p(x_i|\theta_p)] \tag{2}$$

where  $\alpha$  is the frequency of the major locus determined manifestation of the trait. Thus, the likelihood equation for an admixed model is simply the weighted sum of the density functions assumed in the models for each etiologic pattern. The weights are given by the frequencies with which each process occurs. By adding additional frequency

parameters,  $\alpha_j$ , where  $j$  runs over the processes being considered, one can construct more comprehensive, though most likely more difficult to implement, models of a particular trait.

Three aspects of the models just described should be emphasized at this point. First, it is our belief that including a great number of submodels (i.e., assuming a great many processes) in a large model of heterogeneity may not be necessary *if* one is interested in particular major locus determinants of a trait. That is, we hypothesize that in certain situations it will suffice to posit a particular major locus model in combination with a single, more comprehensive (e.g., mixed or polygenic), model since we believe that those pedigrees not compatible with the particular major locus model (i.e., those following other major locus, polygenic, environmental, etc. patterns) may be subsumed under the alternative model. In addition, we also hypothesize that genes with large phenotypic effects may be characterized by pure, single locus submodels in heterogeneity model settings even in cases where background polygenes and complex environmental forces are operative. Obviously, the utility and validity of a such hypotheses will need to be explored in full. Second, the admixed models of the type given in Eq. (2) are not limited to segregation phenomena. Measured genotype [Boerwinkle et al., 1986] and linkage models [Smith, 1963] could be easily incorporated into the framework described above. Third, the current presentation of the admixed model assumes that the pedigrees under scrutiny are sampled randomly from the population at large, as would be the case in an epidemiological study of a trait with an unknown genetic basis. This is unlike the models and approach discussed in Greenberg and Hodge [1985], where it is explicitly assumed that ascertainment on the basis of a disease manifestation of a qualitative trait is employed. It is possible to correct for ascertainment in the case of the admixed models presented here, although, as pointed out by Greenberg and Hodge [1985], such ascertainment may nonetheless yield some bias in the segregation parameter estimates.

### Testing for Heterogeneity

The flexibility of the model type assumed in Eq. (2) does carry with it some potential dangers. For instance, it may be the case that one finds only spurious evidence for a major locus determined manifestation of the trait in question. Such spuriousness could arise as a result of having some statistically spurious "outlier" pedigrees in the sample, or result simply from chance. Thus, it is imperative that one *test* for, or examine the significance of, heterogeneity. The test procedure adopted here derives from the "parametric bootstrap" procedures discussed in Schork and Schork [1989] and Schork [1991d] for other genetic modeling contexts. Bootstrap tests have, in general, an intuitive appeal and many bootstrap tests have excellent asymptotic properties: see, for example, the papers by Beran [1988], Hall and Titterton [1989], Jockel [1986], and Hall and Wilson [1991]. The null hypothesis,  $H_0$ , is taken to be homogeneity of the genetic determinants, and is assumed to be characterized by the unique component models which *collectively* model or characterize the alternative hypothesis,  $H_1$ , of heterogeneity. Thus, if, as in the case of the model assumed in Eq. (2), it is assumed that major locus and polygenic determined manifestations of a trait exist, then the null hypotheses are taken to be characterized by a major locus model *in isolation* and/or a polygene model *in isolation*. Test statistics,  $t$ , are constructed from log-likelihood ratios. For the example just outlined, these statistics would be

$$\begin{aligned}
 t_{H_0:p} &= -2 \log \{L_p(\hat{\theta}_p|X)/L_{m,p}(\hat{\alpha}, \hat{\theta}_m, \hat{\theta}_p|X)\} \\
 t_{H_0:m} &= -2 \log \{L_m(\hat{\theta}_m|X)/L_{m,p}(\hat{\alpha}, \hat{\theta}_m, \hat{\theta}_p|X)\}
 \end{aligned}
 \tag{3}$$

where  $\hat{\theta}$  is an estimate of  $\theta$ .

As noted by many researchers, mixture models often violate the regularity conditions under which  $t$  may be assumed to be asymptotically distributed as  $\chi^2$  [Schork and Schork, 1989]. We consider estimating the distribution of  $t$  and drawing critical values from this estimated distribution. First one computes  $t_{H_0:p}$  or  $t_{H_0:m}$  from the data and obtains parameter estimates  $\hat{\theta}_{p|H_0:p}$ , (or  $\hat{\theta}_{m|H_0:m}$ ) and  $(\hat{\alpha}, \hat{\theta}_m, \hat{\theta}_p)_{H_1}$  (i.e., parameter estimates under the null and alternative hypotheses). One then generates  $r$  data sets  $X_i^*$  where  $i = 1, \dots, r$  from  $H_0$  using  $\hat{\theta}_{p|H_0:p}$  or  $\hat{\theta}_{m|H_0:m}$  as the generating parameter values. For each  $X_i^*, t_{H_0:j}^*$ ,  $j \in \{p, m\}$ , is computed. A critical value for  $t$  taken from the “real” data, with an assumed type I error level of  $\alpha$ , is then given by the  $(1 - \alpha)r$ th order statistic of the  $t_{H_0,i}^*$ . Note that since two tests would be performed in this example (i.e., one for  $H_0 : p$  and one for  $H_0 : m$ ) one might want to correct the alpha error for the test construction appropriately (e.g., by Bonferroni’s method). Significant heterogeneity can be said to exist when all possible homogeneity models (i.e., those taken as components of the heterogeneity model) are rejected. Note that one could also investigate the “reciprocal” test with the associated null hypothesis of heterogeneity, by generating data from  $(\hat{\alpha}, \theta_p, \hat{\theta}_m)_{H_1}$  and computing  $t^*$  for each simulated data set [Schork and Schork, 1989]. The result would give some indication of the plasticity, robustness, and power of the model used to characterize the heterogeneity. Parametric bootstrap tests of the variety just described work well, but can still be greatly improved upon through some analytic adjustments, as shown by Schork [1991e]—see also Beran [1988] for a general discussion of improvements that can be made for some special nested hypothesis testing frameworks. One obvious drawback to the proposed test construction is its heavy simulation load, although this load can be greatly reduced through the use of a parallel computer [Schork, 1991e; Schork and Hardwick, 1992]. The computation times associated with the use of a variety of computers in segregation analysis settings discussed in Schork [1991a,b] can be used to gauge the computational effort of the procedure.

**Classifying Pedigrees**

Once one has determined that significant heterogeneity exists for the trait under scrutiny, it is important to identify individuals or pedigrees compatible with the segregation or environmentally dictated patterns which collectively describe the heterogeneity. The approach to classifying pedigrees advocated in this paper derives from traditional multivariate discrimination methods that minimize the expected cost of misclassification (see, for instance, Johnson and Wichern [1982], pp. 494–497). Simulation methods are used to estimate quantities assumed in the procedure.

Consider a heterogeneity model comprised of  $g$  submodels with associated density functions  $f_i$  ( $i = 1, \dots, g$ ) with parameters  $\theta_i$  and mixing weights  $\alpha_i$ . A pedigree with data vector  $x$  is assigned to a certain pattern, or is said to be compatible with a certain model,  $k$ , on the basis of that  $k$  which minimizes,

$$\sum_{i \neq k}^g \alpha_i f_i(x|\hat{\theta}_i) c(i|k)
 \tag{4}$$

for the pedigree. The  $c(i|k)$  are the "misclassification costs" associated with the assignment of a pedigree compatible with pattern  $k$  to patient  $i$ .  $\hat{\theta}_{i|H_1}$  are parameters associated with submodel  $i$  estimated under the *heterogeneity* hypothesis; as such the  $\hat{\theta}_{i|H_1}$  are *not* parameter estimates for model  $i$  obtained in isolation of the other submodels' parameters assumed in the heterogeneity model.

The  $c(i|k)$  are generally unknown and most likely not reliable when assigned values on the basis of some subjective assessment of pertinent misclassification costs. They should be interpreted as simple "weights" that aid in the proper assignment of pedigrees to patterns. We consider determining optimal  $c(i|k)$  through a simple simulation method. Given estimates  $\hat{\alpha}_i$  of the  $\alpha_i$  and  $\hat{\theta}_{i|H_1}$  of the  $\theta_{i|H_1}$ , one generates data sets *known* to conform to  $H_1$  by using the  $\hat{\alpha}_{i|H_1}$  and  $\hat{\theta}_{i|H_1}$  as generating values. That is, one generates  $\hat{\alpha}N$  pedigrees conforming to  $f_1(x|\hat{\theta}_{1|H_1})$ ,  $\hat{\alpha}_2N$  conforming to  $f_2(x|\hat{\theta}_{2|H_1})$ , etc. One can then determine the  $\hat{c}(i|k)$  which minimize some overall measure of misclassification (e.g., the sum of misclassifications) by comparing those classifications made using Eq. (4) and the exact classifications known from the simulated data. This procedure can be implemented most easily by simply investigating a grid of various  $\hat{c}(i|k)$  values using some uniform increment. The resulting  $\hat{c}(i|k)$  which meet the overall misclassification criterion are then used with Eq. (4) to classify the "real" pedigree data.

Three aspects of the proposed procedure should be emphasized. First, the proper estimation of the  $\hat{c}(i|k)$  is crucial. The submodels of a heterogeneity model will typically be non-nested with respect to one another and hence be on radically different scales. Thus, one could not compare the nonweighted values of, say,  $f_i(x|\hat{\theta}_i)$  and  $f_j(x|\hat{\theta}_j)$  [or  $\hat{\alpha} f_i(x|\hat{\theta}_i)$  and  $\hat{\alpha}_j f_j(x|\hat{\theta}_j)$ ] for that matter] to make assignments since  $f_j$  may actually be much *smaller* than  $f_i$  for a particular pedigree but still actually be the appropriate model for the pedigree. This is unlike most traditional discrimination or classification settings where the models typically have the same distributional form (e.g., multivariate normal) but differ only with respect to their parameter values and are hence on a comparable scale. However, in linkage and other settings, where one assumes that only a recombination fraction or other single parameter varies among subgroups in the population at large, then nested-model based procedures such as  $\chi^2$ -based likelihood ratios can be used, as described in Smith [1963]. Second, one may want to "tune" the overall measure of misclassification used in the simulation procedure to account for possible imbalances in misclassification rates. For instance, if minimization of the (simple) sum of misclassifications is used as the criterion for determining the  $\hat{c}(i|k)$  then it may be that the resulting classification rule will have a greater tendency to misclassify pedigrees in a certain direction (e.g., it may assign pedigrees to a major locus pattern that truly follow a polygene pattern but not vice versa). As a result, one may want to add constraints on the misclassification criterion, such as having the misclassifications equally distributed among the submodel types, or by not allowing misassignments to a major locus pattern to occur. Third, since it is rare that one collects pedigrees with the same size and structure, the simulation procedure for determining the  $\hat{c}(i|k)$  should be repeated a few times with the pedigree sizes and structures generated under each submodel setting allowed to vary with the real pedigree sizes and structures. The average  $\hat{c}(i|k)$  can then be used for the real data.

## RESULTS

In order to display and assess the reliability of the proposed methods, two simulation studies investigating the reliability of the parameter estimation process in segrega-

tion analysis settings, a simulation study investigating the classification procedure, and a simulation study describing a situation involving measured genotype information, were employed. In addition, the methodology as applied in linkage situations is showcased with some data relevant to hypertension research taken from the Tecumseh Community Health and Tecumseh Blood Pressure studies.

### Segregation Analysis

To investigate the parameter estimation process we assumed a trait with 2-allele single locus and polygenic manifestations. It was further assumed that these manifestations were equally prevalent (i.e.,  $\alpha_1 = \alpha_2 = 1/2$ ). For the major locus, an allele frequency,  $p_A$ , of 0.316, mean genotype affects of  $\mu_{AA} = 12.0$ ,  $\mu_{Aa} = 10.0$ , and  $\mu_{aa} = 9$ , and variance of  $\sigma_{pe}^2 = 1.0$  were assumed. Thus, the “high” genotype with large effects,  $AA$ , had a frequency of 0.1. The polygene manifestation was such that an overall mean of  $\mu_p = 10.0$ , additive variance  $\sigma_a^2 = 0.75$ , dominance variance  $\sigma_d^2 = 0.25$ , and environmental variance  $\sigma_{pe}^2 = 1.0$  were assumed. Though somewhat artificial, this heterogeneous setting is sufficiently challenging and realistic enough (e.g., because of the relatively low “high” genotype frequency, small major locus mean genotype effect separation, and low to moderate heritability, 0.375, for the polygenic manifestation of the trait) to be computationally difficult and capable of gauging the usefulness of the model parameter estimation procedure. Twenty-four pedigrees of the size and structure displayed in Figure 1 were generated (i.e., 12 pedigrees compatible with the major locus setting and 12 compatible with the polygenic setting). Estimates of the parameters were obtained by maximizing the appropriate likelihood equation as in Eq. (2) over the simulated data using the NPSOL package [Gill et al., 1984] and relevant subroutines written by the author (available upon request through Internet e-mail: Nicholas\_Schork@um.cc.umich.edu). This simulation/estimation process was repeated 100 times, and was also performed with 36, 48, and 60 pedigrees. Mean parameter estimates and standard errors obtained over the 100 simulations for each sample size are displayed in Table I along with average mean squared errors calculated by considering the sum of squared distance of the parameter estimates from the

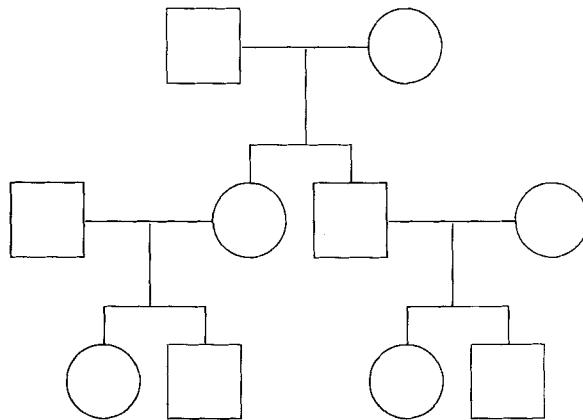


Fig. 1. Pedigree structure used in the simulation study investigating the use of admixed models in segregation analysis settings.

**TABLE I. Average Parameter Estimates and Mean Squared Errors Obtained Over 100 Simulations Involving a Heterogeneity Model Assuming Major Locus and Polygenic Manifestations of a Quantitative Trait, for Different Numbers of 10 Member Pedigrees, With Generating Values,  $gv^*$**

par	$gv$	24		36		48		60	
$\alpha$	0.500	0.516	0.026	0.555	0.026	0.499	0.029	0.530	0.026
ML									
$p(A)$	0.316	0.356	0.019	0.364	0.014	0.353	0.013	0.365	0.013
$u(AA)$	12.00	11.88	0.084	12.01	0.081	12.13	0.089	11.98	0.067
$u(Aa)$	10.00	9.93	0.061	9.95	0.047	10.00	0.056	9.94	0.045
$u(aa)$	9.00	8.94	0.046	8.95	0.042	8.98	0.037	8.92	0.040
$v(e)$	1.00	0.897	0.091	0.856	0.028	0.814	0.029	0.924	0.032
Poly									
$u$	10.00	10.00	0.043	9.98	0.048	10.01	0.041	9.94	0.038
$v(a)$	0.750	0.440	0.047	0.579	0.064	0.654	0.061	0.556	0.049
$v(d)$	0.250	0.577	0.067	0.447	0.063	0.418	0.055	0.341	0.050
$v(e)$	1.00	0.844	0.071	0.957	0.091	0.858	0.060	0.938	0.068
MSE		0.312	0.022	0.312	0.026	0.267	0.022	0.213	0.018
$\alpha$ MSE		0.065	0.006	0.071	0.007	0.080	0.007	0.068	0.007
ML MSE		0.300	0.038	0.233	0.028	0.271	0.036	0.186	0.020
Poly MSE		0.345	0.035	0.363	0.045	0.287	0.047	0.282	0.043

\*par, parameter;  $\alpha$ , mixing proportion;  $p(A)$  = frequency of the major locus  $A$  allele;  $u(-)$ , mean of genotype ( $-$ );  $v(-)$  variance associated with environment (e), additive genetic (a), and dominance genetic (d) effects; ML, major locus submodel; Poly, polygenic submodel; MSE, mean squared error;  $\alpha$ , ML, Poly MSE, MSE associated with  $\alpha$ , ML, and Poly parameters. Note: for parameters the first number is the average, the second the MSE, for the MSEs the first is the average, the second the standard error.

generating values for each simulation experiment. Table I suggests that parameter estimation is fairly reliable in this setting and generally improves, as expected, with increases in sample size. It is interesting to note from Table I that major locus parameters are estimated more reliably than the polygene parameters. This is encouraging since it is assumed that major locus settings are of primary interest because of their amenability to current laboratory marker-based linkage mapping strategies.

A simulation study was also performed in order to investigate how well major locus submodel parameters can be estimated in the presence of polygenic effects acting in concert with the major locus effects. In this way a brief exploration of one of the hypotheses concerning the inclusion of "pure" single locus submodels in the heterogeneity models put forth in the Methods section could be pursued. One hundred sets of 36 pedigrees of the Figure 1 variety following the parameter configuration used in the first simulation study were generated, except that polygenic effects that would amount to residual heritabilities of 0.1, 0.25, and 0.5 were imposed on the major locus data. Table II shows mean parameter estimates, standard errors, and MSEs obtained over the 100 replicate runs at each setting (note: column 2 of Table I is included as a reference). Table II suggests that the bias in the major locus parameter estimation is not excessive in the presence of polygenic background "noise," though, as expected, there is greater variability in the estimates.

### Segregation Model Classification

The heterogeneity patterns and models used in the construction of Table I with 36 pedigrees of the Figure 1 variety were assumed, so that determination of the optimal weights used in Eq. (4) involved the comparison of the densities assumed in the mod-



TABLE II. Average Parameter Estimates and Mean Squared Errors Obtained Over 100 Simulations Involving a Heterogeneity Model Assuming Major Locus and Polygenic Manifestations of a Quantitative Trait, for 36 10-Member Pedigrees, in Which Polygenic Effects Have Been Induced on the Major Locus Manifestation\*

par	gv	$H = 0$		$H = 0.1$		$H = 0.25$		$H = 0.5$	
$\alpha$	0.500	0.555	0.026	0.512	0.028	0.482	0.028	0.403	0.027
ML									
$p(A)$	0.316	0.364	0.014	0.352	0.015	0.347	0.015	0.375	0.017
$u(AA)$	12.00	12.01	0.081	12.12	0.089	12.24	0.096	12.08	0.128
$u(Aa)$	10.00	9.95	0.047	10.04	0.054	10.07	0.057	10.09	0.067
$u(aa)$	9.00	8.95	0.042	8.81	0.041	8.81	0.050	8.67	0.070
$v(e)$	1.00	0.858	0.028	0.829	0.037	0.855	0.041	0.881	0.045
Poly									
$u$	10.00	9.98	0.048	9.96	0.050	9.96	0.071	9.90	0.047
$v(a)$	0.750	0.579	0.064	0.572	0.054	0.596	0.064	0.761	0.054
$v(d)$	0.250	0.447	0.063	0.513	0.068	0.544	0.071	0.445	0.064
$v(e)$	1.00	0.957	0.091	0.799	0.064	0.752	0.064	0.876	0.065
MSE		0.312	0.026	0.302	0.020	0.340	0.026	0.432	0.034
$\alpha$ MSE		0.071	0.007	0.074	0.007	0.079	0.067	0.078	0.006
ML MSE		0.233	0.028	0.285	0.028	0.352	0.043	0.572	0.066
Poly MSE		0.363	0.045	0.381	0.043	0.391	0.047	0.345	0.044

\*See Table I;  $H$ , residual polygenic heritability added over the major locus manifestation of the trait.

els for major locus and polygenic manifestations of the trait. Since only two submodels were involved, classification to the major locus pattern was based on whether or not the following inequality was obtained for a particular pedigree with data vector  $x_i$ :

$$\alpha f_m(x_i | \theta_{m|H_1}) c(m|p) > (1 - \alpha) f_p(x_i | \theta_{p|H_1}) c(p|m). \tag{5}$$

If (5) did not hold then the pedigree was assumed to be consistent with a polygenic pattern. Twenty-five data sets were generated for which optimal  $c(i|k)$  were estimated. It was assumed that the misclassifications resulting in an erroneous assignment of a “polygenic pedigree” as being consistent with a major locus pedigree should be minimized along with the total number of misclassifications. Thus, the misclassification criterion for each of the 25 data sets involved minimization of the sum of polygene misassignments and the total number of misassignments (i.e., 2 times the number of polygene to major locus misassignments plus 1 times the number of major locus to polygene misassignments). Since comparison of only two submodels was required, only one weight [here taken to be  $c(m|p)$ ] needed to be obtained as the other [here taken to be  $c(p|m)$ ] could be set to 1.

Optimal  $c(m|p)$  were determined by reading off values associated with the minimum misclassification value over a grid of 1000  $c(m|p)$  values uniformly dispersed over a range (0 to 5) of  $c(m|p)$  values. Table III displays the results obtained for each of the 25 data sets. The average  $c(m|p)$  over the 25 runs was 0.40. Thus, if a real data set produced parameter estimates consistent with the generating values used in Table I over 36 pedigrees of the appropriate size, classification of the pedigrees would involve  $\hat{c}(m|p) = 0.4$  and  $\hat{c}(p|m) = 1.0$  with Eq. (5). Note that although these weights may rarely assign a “polygene pedigree” a major locus pattern, they do so at a cost of misclassifying a large number of “major locus pedigrees”: from Table III it can be

**Table III. Number of Misclassifications and Weights Obtained From 25 Data Sets of 36 10-Member Pedigrees Each Known to Conform to the Admixed Model Setting Outlined in Table I, Based on Minimization of the Total Number of Misclassifications With the Constraint of Obtaining the Fewest Misassignments to a Major Locus Pattern; Each Cell of 4 Numbers Contains Results From One of the Simulations\***

8	1	6	1	11	2	12	1	7	3
9	0.37	7	2.0	10	2.0	13	0.02	10	0.02
16	0	18	0	13	0	18	0	15	0
14	0.03	17	0.00	13	1.30	16	0.00	13	0.01
15	0	15	0	4	4	18	0	7	9
15	0.11	13	0.06	8	0.02	18	0.00	13	1.00
7	5	18	0	15	0	11	1	11	2
12	0.31	17	0.00	15	0.12	10	0.55	13	0.08
18	0	15	1	18	0	11	1	18	0
17	0.00	15	0.03	11	0.00	9	1.50	16	0.00

\*Upper right entry, number of major locus to polygene misclassifications; upper left, number of polygene to major locus misclassifications; lower right, total number of misclassifications if the criterion was simply the total number of misclassifications; lower left, optimal  $c(m|p)$  value.

seen that whereas on average only 0.88 polygenic pedigrees were misclassified, some 13.4 or 74% of the major locus pedigrees were, on average, misclassified. This low sensitivity may not be too serious a problem, however, if one wants to cull out the "unequivocal" major locus type pedigrees for, say, a linkage or in-depth physiological study.

It is also somewhat disconcerting that the optimal  $c(m|p)$  varied widely through the 25 replicate runs. However, additional information gleaned from the simulations required for the classification method can provide further insight into choosing the optimal  $c(i|k)$ . A significant relationship between how close the estimated parameters were to their generating values (measured by mean squared error, MSE) and the number of misclassifications was found ( $p < .05$  by rank correlation) in the simulation data. Thus, one could weight the  $c(i|k)$  by how reliably the parameters were estimated for each of the simulations to create the  $c(i|k)$  to be used with the real data.

### Measured Genotype Analysis

To investigate the utility of admixed models in measured genotype settings of the type described in Boerwinkle et al. [1986] and George and Elston [1987], 100 nuclear families of size 5 (i.e., each with three offspring) were generated. It was assumed that  $1/2$  of these families had a trait that was affected by a locus with 2 alleles,  $A$  and  $a$ , with frequencies  $p_A = p_a = 0.5$ , and mean genotype effects of  $\mu_{AA} = 8$ ,  $\mu_{Aa} = 10$  and  $\mu_{aa} = 12$ . It was also assumed that operating in the background of this locus were polygenic effects that produced a residual heritability of 0.5 (where  $\sigma_a^2 = \sigma_e^2 = 0.5$ ). Families whose traits were not affected by this locus were assumed to have phenotypes that could be characterized by a mean effect of 10, and additive and environmental variances of 1.0.

Parameter estimates were obtained from these data using the NPSOL package for the following models: a "pure" polygene model, a pure measured genotype model, and a heterogeneity model comprised of pure polygene and measured genotype models. Results of the model fitting are displayed in the first 3 rows of Table IV. In order to investigate the significance of the heterogeneity, 99 data sets of 100 nuclear fami-

**TABLE IV. Results of Model Fitting Involving Data Derived From a Mixture of Nuclear Families Following Genotype Specific and Polygene Influenced Trait Patterns; the Lower Panels Display Parameter Estimates Associated With Data Generated to Test Hypotheses About the Data in the Upper Panel\***

Data/model	$\alpha$	m-AA	m-Aa	m-aa	add-v	env-v	m-p	a-vp	e-vp	log-Like
Original data										
Generating	0.500	8.000	10.00	12.00	0.500	0.500	10.00	1.000	1.000	
Polygene							10.10	2.183	0.329	-1254.23
Measured genotype										-1194.78
Admixed	0.545	8.173	10.00	12.12	0.792	0.226	10.17	1.674	0.299	-1160.98
Polygene data										
Polygene							10.09	2.183	0.329	-1706.31
Measured genotype							(0.016)	(0.012)	(0.011)	(2.313)
Admixed	0.814	10.04	10.07	10.07	2.038	0.316	9.144	0.983	0.409	-1705.20
	(0.027)	(0.039)	(0.031)	(0.035)	(0.052)	(0.079)	(0.071)	(0.139)	(0.077)	(2.567)
Measured genotype data										
Polygene							10.14	2.265	0.276	-1701.45
Measured genotype							(0.010)	(0.024)	(0.010)	(2.811)
Admixed	0.983	9.177	10.02	11.32	1.710	0.256	8.070	0.393	0.178	-1620.71
	(0.003)	(0.013)	(0.010)	(0.014)	(0.023)	(0.010)	(1.048)	(0.095)	(0.045)	(2.666)

\* $\alpha$ , mixing weight; m-AA, Aa, aa, mean genotype effects for the measured locus; v-a,e = additive and environmental variance; m-p = mean of the polygenic determined subgroup. For the simulated polygenic and measured genotype data, the upper entry is the mean and the lower entry the standard error over the 99 replications (in parentheses).

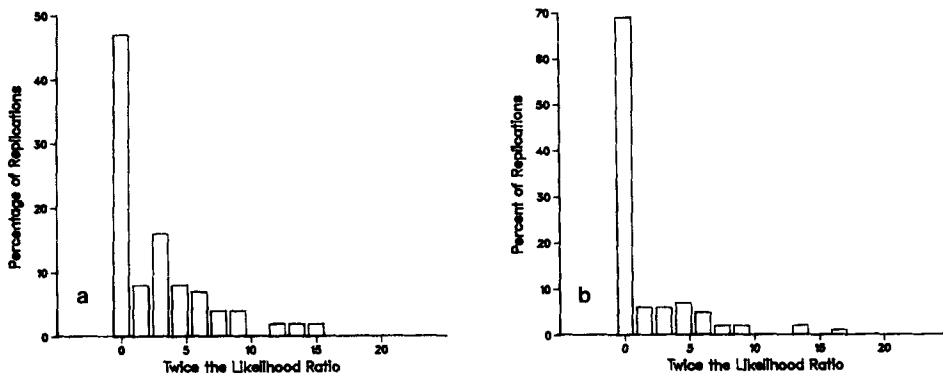


Fig. 2. (a) Simulated null distribution of twice the log-likelihood ratio (LR) statistic implicating polygene and heterogeneity models. The LR statistic obtained from the "original" data was 186.5. (b) Simulated null distribution of twice the LR statistic implicating measured genotype and heterogeneity models. The LR statistic obtained from the original data was 67.6.

lies of size 5 were generated using parameter estimates of both the pure polygene and measured genotype models obtained from the original data set (i.e., those parameter values listed in the first 2 rows of Table IV). Average parameter estimates and associated standard errors obtained from each setting over the  $2 \times 99$  data sets are given in the bottom 6 rows of Table IV. Likelihood ratios (LRs) from each replicate run were obtained to compare models. Figure 2a and b displays the distributions of the LRs comparing the heterogeneity and pure polygene and measured genotype models, respectively, using the 99 relevant data sets for each. It can be seen from Figure 2a and b that the LRs from the original data greatly exceed the LRs obtained from the 99 pure polygene and 99 pure measured genotype data sets. Thus, the homogeneity hypothesis can be rejected for the original data. A great many of the pure polygene and pure measured genotype data sets produced estimates of 1.0 for the mixing weight for the admixed model, resulting in the "automatic rejection" of the heterogeneity hypothesis and a log-LR value of 0.0, as is evidenced in Figure 2.

### Sib-Pair Linkage Analysis

To demonstrate the application of admixed models in linkage analysis settings, some results obtained from a study of the relationship between systolic blood pressure (SBP) and the MN locus using data gathered in the Tecumseh Community Health and Tecumseh Blood Pressure studies are described. A manuscript describing the full details of the study is forthcoming (contact the author for further details). A total of 4,621 sib-pairs derived from 11,563 participants in the Tecumseh studies were available for study. The average number of *sib-pairs* per family was 1.2. A bootstrap-based regression procedure that made use of only independent (i.e., nongenetically related or from the same pedigree) participants at each iteration was used to derive a linear model that was used, in turn, to adjust SBP values for sex, age, weight, height, and possible interaction terms. The differences in adjusted SBP values for sibs were then investigated for possible heterogeneous linkage relationships between a postulated SBP locus and the MN locus using extensions of the two sib-pair linkage strategies outlined in Haseman and Elston [1972].

The first procedure involved the regression of squared sib-pair adjusted SBP differences on the number of alleles shared by the sibs that are identical by descent. This procedure is outlined in Sections II to IV of Haseman and Elston [1972], and posits 3 parameters: an intercept term, a slope, and a residual variance term. A negative slope is indicative of linkage. A heterogeneity model assuming the basic regression model is formed by a mixture of such regressions, with each component in the mixture having its own associated intercept, slope, and variance terms. In order to compute likelihoods a parametric form for the error distributions of the regressions needed to be assumed. The distribution associated with the square of a normally distributed variable was used, although analyses involving other distributional forms yielded surprisingly similar results.

The second procedure made use of the likelihood-based mixture distribution model outlined in Section V of Haseman and Elston [1972]. This model is parameterized by a recombination fraction, an allele frequency associated with the hypothesized trait locus (marker allele frequencies are assumed known), an additive effect parameter, a dominance effect parameter, and a residual variance parameter. For further details one should consult Haseman and Elston [1972] and Blackwelder and Elston [1974] for some corrections and clarifications. A heterogeneity model assuming this basic model can be constructed from a simple mixture of such models.

Although it can be shown that the use of nonindependent sib-pairs in sib-pair linkage analysis settings has a minimal effect on the interpretation of the results [Blackwelder and Elston, 1982; RC Elston, personal communication], it is not clear what effects the use of nonindependent sib-pairs would have on the mixing weight and other parameters in a linkage *heterogeneity* model setting. As such, the following analytic strategy was adopted. One sib-pair from each pedigree was selected at random and used to estimate linkage parameters. This process was repeated 100 times. Mean parameter estimates and standard errors obtained over the 100 replicate runs were used to draw inferences about possible linkage relationships. Table V displays results. It can be seen from Table V that although a subgroup was found using the regression procedure that presents a negative slope, the value of the slope in this group was not significantly different from 0. However, the mixture distribution heterogeneity model presents a subgroup of a similar size to that obtained with the regression procedure with a recombination fraction that is significantly different from 0.5. In addition, it is fairly clear from Table V that two distinct subgroups exist in terms of adjusted blood pressure difference (that is, without reference to the MN locus)—a finding that is of considerable interest in its own right. Further analysis of these data have shown evidence of sex-specificity in the linkage relationship, which suggests a role for general heterogeneity modeling as a means of screening for sources of variation not ubiquitous in the population at large but still important etiologically.

## DISCUSSION

The models and methods outlined in this paper have a great potential to supplement and/or complement traditional methods used to analyze human quantitative traits. Heterogeneity can easily obscure the detection of relevant genetic effects if not accounted for properly. In addition, heterogeneity is more than likely a common enough phenomenon to warrant attention and consideration for many quantitative traits with disease manifestations, such as blood pressure and cholesterol.

**TABLE V. Sib-Pair Linkage Results Using Two Different Sib-Pair Linkage Techniques on 100 Runs Using Independent Sib-Pairs From the Tecumseh Studies for Adjusted Systolic Blood Pressure Values and the MN Locus<sup>†</sup>**

Parameter	$\alpha$	Regression technique:						var					
		y-int	slope	var	y-int	slope	var						
Homogeneity model		0.761* (0.002)	-0.001 (0.001)	0.395* (0.002)									
Heterogeneity model	0.509* (0.008)	0.490* (0.117)	-0.073 (0.058)	0.217* (0.066)	1.010* (0.040)	0.151* (0.015)	0.315* (0.016)						
Parameter	$\alpha$	Mixture distribution technique						var					
		rec	freq	add	dom	var	rec		freq	add	dom	var	
Homogeneity model		0.444* (0.008)	0.436* (0.030)	0.952* (0.045)	-0.194* (0.096)	0.551* (0.012)							
Heterogeneity model	0.468* (0.026)	0.179* (0.018)	0.508* (0.026)	0.517* (0.052)	0.160 (0.080)	0.614* (0.056)	0.444* (0.009)	0.466* (0.027)	0.945* (0.067)	-0.072 (0.098)	0.707* (0.042)		

<sup>†</sup> $\alpha$ , mixing weight; y-int, y-intercept; var, residual variance; rec, recombination fraction; freq, trait locus allele frequency; add, additive effect; dom, dominance effect. The upper entry for each model is the mean and the lower entry the standard error over the 100 replications (in parentheses).

\*Significance from boundary values based on the standard error at the 5% level.

The proposed models also have the potential to offer more than just a characterization of the heterogeneity of a particular trait, since one could also use them to identify pedigrees or individuals showing a pattern consistent with a particular etiologic or genetic pattern as a precursor to a more in-depth investigation of a particular pattern. This approach, of identifying pedigrees most compatible with a certain segregation pattern, is not entirely new. Beaty [1980], Williams and Lalouel [1982], Burns [1982], Lalouel et al. [1983], Boehnke and Lange [1984], Moll et al. [1984, 1989], Beaty and Boughman [1986], and Rebbeck et al. [1991] all consider the problem of isolating pedigrees most consistent with a particular genetic hypothesis by fitting two competing models to an entire data set and then more or less examining the likelihood that each pedigree is compatible with one or the other hypothesis. This strategy is intuitive if the goal is to identify pedigrees showing the clearest compatibility with a particular segregation pattern—as would be beneficial as a precursor to a linkage study of the trait, as was explicitly assumed by Boehnke and Moll [1989] for the case of a rare dominant quantitative disorder. However, this strategy may be problematic if the goal is to investigate possible genetic heterogeneity or to identify families showing supposedly *divergent* etiologic patterns for a trait. This is so because with this (traditional) approach the parameters used to characterize each genetic or environmental model are estimated from the *entire* data set with no correction for the model misspecification that arises in the case of those families not consistent with the particular model whose parameters are being estimated at the time. In fairness to this approach, however, it should be emphasized that no investigation of the bias (in model selection or parameter estimation) has been undertaken in cases where it is known that some pedigrees do not conform to a certain model even though that model characterizes other pedigrees well. Presumably, however, the bias could be quite large. In this light, the methodology elaborated in this paper can be seen as offering a way to account for such misspecification in a meaningful way.

Despite the intuitive appeal and utility of the proposed models, more work on them is desperately needed if they are to be used routinely. For instance, power studies investigating how easy or difficult it is to detect certain segregation or linkage phenomena in the presence of others are in order. In addition, it is important to determine how many families of a specific type (e.g., those following a rare recessive gene pattern of inheritance) are needed to detect heterogeneity when the relevant model type is combined with more general models, such as the mixed model or a pure polygene model. Ultimately, however, this paper offers a *step* in a direction not yet taken in the analysis of many traits for which such a direction may be appropriate.

## REFERENCES

- Beaty TH (1980): Discriminating among single locus models using small pedigrees. *Am J Med Genet* 6:229–240.
- Beaty TH, Boughman JA (1986): Problems in detecting etiologic heterogeneity in genetic disease illustrated with retinitis pigmentosa. *Am J Med Genet* 24:493–504.
- Beran R (1988): Prepivoting test statistics: A bootstrap view of asymptotic refinements. *JASA* 83:687–697.
- Blackwelder WC, Elston RC (1974): Comment on Dr. Robertson's Communication (letter to the editor). *Behav Genet* 4:97–99.
- Blackwelder WC, Elston RC (1982): Power and robustness of sib-pair linkage tests and extensions to larger sibships. *Commun Statist Theory Methods* 11:449–484.

- Boehnke M, Lange K (1984): Ascertainment and goodness of fit of variance component models for pedigree data. In Rao DC, Elston RC, Kuller LH, Fenleib M, Carter C, Havlik R (eds): "Genetic Epidemiology of Coronary Heart Disease: Past, Present, and Future." New York: Alan R. Liss, 173–192.
- Boehnke M, Moll PP (1989): Identifying pedigrees segregating at a major locus for a quantitative trait: an efficient strategy for linkage analysis. *Am J Hum Genet* 44:216–224.
- Boerwinkle E, Chakraborty R, Sing CF (1986): The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann Hum Genet* 50:181–194.
- Bonney GE (1984): On the statistical determination of major gene mechanisms in continuous human traits: Regressive models. *Am J Med Genet* 35:816–826.
- Burns TL (1982): Sampling considerations for the determination of genetic transmission mechanisms in quantitative traits. Ph.D. thesis, University of Michigan, Ann Arbor.
- Demenaïs FM, Murrigande C, Bonney GE (1990): Search for faster methods of fitting the regressive models to quantitative traits. *Genet Epidemiol* 7:319–334.
- Elston RC (1981): Segregation analysis. In Harris H, Hirshhorn K (eds): "Advances in Human Genetics." New York: Plenum, 63–120.
- Elston RC, Namboodiri KK, Glueck CJ, Fallat R, Tsang R, Lueba V (1975): A study of the genetic transmission of hypercholesteremia and hypertriglyceridemia in a 195-member kindred. *Ann Hum Genet* 39:67–87.
- George VT, Elston RC (1987): Testing the association between polymorphic markers and quantitative traits in pedigrees. *Genet Epidemiol* 4:193–201.
- Gill PE, Murray W, Saunders MA, Wright MA (1984): User's guide for SOL/NPSOL: A FORTRAN package for non-linear programming. Stanford University, Department of Operations Research, Report SOL 83-1.
- Greenberg DA, Hodge SE (1985): The heterogeneity problem. I. Separating genetic from environmental forms of the same disease. *Am J Med Genet* 21:357–371.
- Hall P, Titterington DM (1989): The effect of simulation order on level accuracy and power of Monte Carlo tests. *JR Statist Soc B* 51:459–467.
- Hall P, Wilson SR (1991): Two guidelines for bootstrap hypothesis testing. *Biometrics* 47:757–762.
- Haseman JK, Elston RC (1972): The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:1–19.
- Hasstedt SJ (1982): A mixed model likelihood approximation on large pedigrees. *Comput Biomed Res* 15:295–307.
- Jockel KH (1986): Finite sample properties and asymptotic efficiency of Monte Carlo tests. *Ann Statist* 14:336–347.
- Johnson RA, Wichern DW (1982): "Applied Multivariate Statistical Analysis." Englewood Cliffs, NJ: Prentice-Hall.
- Lalouel JM, Darlu P, Henrotte JG, Rao DC (1983): Genetic regulation of plasma and red blood cell magnesium concentrations in man. II. Segregation analysis. *Am J Hum Genet* 35:938–950.
- Lange K, Westlake J, Spence MA (1976): Extensions to pedigree analysis III. Variance components by the scoring method. *Ann Hum Genet* 39:485–491.
- Moll PP, Berry TD, Weidman WH, Ellefson R, Gordon H, Kottke BA (1984): Detection of genetic heterogeneity among pedigrees through complex segregation analysis: An application to hypercholesteremia. *Am J Hum Genet* 36:197–211.
- Moll PP, Michaels VV, Weidman WH, Kottke BA (1989): Genetic determination of plasma apolipoprotein AI in a population-based sample. *Am J Hum Genet* 44: 124–139.
- Ott J (1977): Counting methods (EM algorithm) in human pedigree analysis: Linkage and segregation analysis. *Ann Hum Genet* 40:443–454.
- Ott J (1979): Maximum likelihood estimation by counting methods under polygenic and mixed models in human pedigree analysis. *Am J Hum Genet* 21:161–175.
- Pearson K (1893): Contributions to the mathematical theory of evolution. *Phil Transact A* 185:71–110.
- Pearson K (1895): Contributions to the mathematical theory of evolution. V. Skew variation in homogeneous material. *Phil Transact A* 186:342–414.
- Rebeck TR, Turner ST, Michels VV, Moll PP (1991): Genetic and environmental explanations for the distribution of sodium-lithium countertransport in pedigrees from Rochester, MN. *Am J Hum Genet* 48:1092–1104.
- Risch N, Baron M (1982): X-linkage and genetic heterogeneity in bipolar related major affective illness: Reanalysis of linkage data. *Ann Hum Genet* 46:153–166.



- Schork NJ (1991a): Extended pedigree patterned covariance matrix mixed models for quantitative phenotype analysis. *Genet Epidemiol*, in press.
- Schork NJ (1991b): Efficient computation of patterned covariance mixed models in quantitative segregation analysis. *Genet Epidemiol* 8:29–46.
- Schork NJ (1991c): The parallel computation of pedigree likelihoods. Proceedings of the 23rd Symposium on the Interface of Computing Science and Statistics, Fairfax Station, Virginia: Interface Foundation of North America, Inc., 262–265.
- Schork NJ (1991d): Bootstrapping likelihood ratios in quantitative genetics. In LePage R, Billard L (eds): “Exploring the Limits of the Bootstrap.” New York: John Wiley, 389–396.
- Schork NJ (1991e): Bootstrap tests of non-nested hypotheses (abstract). *Biomet Bull* 8:22.
- Schork NJ, Schork MA (1989): Testing separate families of segregation hypotheses: Bootstrap methods. *Am J Hum Genet* 45:803–813.
- Schork NJ, Schork MA (1991): The relative efficiency and power of small pedigree studies of the heritability of a quantitative trait. *Hum Hered*, in press.
- Schork NJ, Hardwick JP (1992): Supercomputer intensive multivariable randomization tests. Proceedings of the 22nd Symposium on the Interface of Computing Science and Statistics, New York: Springer-Verlag, 509–513.
- Schork NJ, Weder AB, Schork MA, Bassett Dr, Julius S (1990a): Disease entities, mixed multinormal distributions, and the role of the hyperkinetic state in the pathophysiology of hypertension. *Statist Med* 9:301–314.
- Schork NJ, Weder AB, Schork MA (1990b): On the asymmetry of biological frequency distributions. *Genet Epidemiol* 7:427–446.
- Smith CAB (1963): Testing for heterogeneity of recombination fraction values in human genetics. *Ann Hum Genet* 27:175–182.
- Weder AB, Schork NJ (1989): Mixture analysis of erythrocyte sodium-lithium countertransport and blood pressure. *Hypertension* 13:145–150.
- Weder AB, Schork NJ, Julius S. (1991): Linkage of MN Locus with erythrocyte lithium-sodium countertransport in Tecumseh, Michigan. *Hypertension* 17:997–981.
- Williams WR, Laouel JM (1982): Complex segregation analysis of hyperlipidemia in a Seattle sample. *Hum Hered* 32:24–36.

**Edited by Aravinda Chakravarti**