

Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. V. Variance components estimated from pedigrees

By PATRICIA P. MOLL, RACHEL POWSNER AND CHARLES F. SING

*Department of Human Genetics, University of Michigan Medical School,
Ann Arbor, Michigan 48109*

INTRODUCTION

The human geneticist who wishes to estimate the genetic and environmental components of a quantitative trait has the choice of several sampling designs. Sets of monozygous and dizygous twins, parents and their offspring, and full siblings are often the most convenient samples but invariably yield estimators of genetic variance which are biased by shared environmental factors. More recently, fixed combinations of related and unrelated individuals have been proposed as sampling strategies to improve the separation of genetic and shared environmental correlations between relatives. These fixed designs take the form of family sets (Schull *et al.* 1970; Chakraborty *et al.* 1977) or sets of monozygous twins, their spouses and offspring (Nance & Corey, 1976). Unfortunately, human data cannot easily be collected according to a fixed design. A convenient sample usually consists of families which vary in the genetic relatedness of their members.

Most human studies have utilized regression coefficients and interclass correlations estimated from pairs of individuals or intraclass correlations from subsets of individuals selected from a random sample of families. These statistics are generally contrasted to estimate the fractions of phenotypic variance due to genetic and environmental factors. The strategies available for combining these estimators do not adequately account for the double counting of individuals and alleles which may occur. For example, the full sib correlations and the parent-offspring correlations are usually estimated from individuals drawn from the same array of nuclear families. Elston (1975) has shown that the correlations between these correlation estimates may not be trivial.

Despite the obvious implication, the precise effect of using correlated estimates on inferences about heritability estimates is not clear at this time. In order to minimize the inflation of type II error it is not uncommon to discard a large portion of the data. For example, in the Tecumseh Community Health Study, from the 6366 individuals typed for 12 blood markers, less than 100 genetically independent family sets, consisting of an index, a sib and a cousin could be constructed (Orr, personal communication). If the assumption of independence of family sets was rigorously met, only 5% of the available data could be utilized and the effect of the non-independent index-sib and index-cousin correlations would still be unresolved. More importantly, samples of pairs or sets of relatives are usually not randomly sampled and therefore may not represent the same phenotypic variability of a quantitative trait that is determined by the frequency distribution of genotypes and environments among and within pedigrees which define the population.

Alternatively, we may use the procedure of maximum likelihood estimation (MLE) to combine all of the information available in a sample of randomly selected pedigrees to obtain

estimates of the sources of phenotypic variability while accounting for the correlations between pairs of observations. A likelihood model which makes this strategy feasible has been proposed by Elston & Stewart (1971) and extended by Lange, Westlake & Spence (1976). It combines phenotypic, genealogical and environmental information to estimate both genetic and environmental components of variance from pedigrees.

We have analysed the pedigree data from Tecumseh, Michigan, to estimate genetic and environmental components of the normal variability in non-fasting total serum cholesterol. The first objective of this study was to compare the maximum likelihood estimates of genetic, household environment, and special (random) environmental effects with estimates obtained from a least squares analysis of correlations between sibs, parent-offspring, spouses and cousins from the same population (Sing & Orr, 1978). Our primary interest here was to determine whether different methods of analysis will lead to substantially different conclusions when applied to the same data set. Our second objective was to determine the effects of non-representative sampling of pedigree sizes on inferences about variance components when the MLE procedure is used.

METHODS

The model

A linear biological model which might be expected to correspond reasonably closely with reality is the following, in which the phenotypic variance (σ_p^2) for unrelated individuals may be written:

$$\sigma_p^2 = \sigma_a^2 + \sigma_d^2 + \sigma_{Ce}^2 + \sigma_{Se}^2, \quad (1)$$

where σ_a^2 is the additive genetic variance, σ_d^2 is the variance of deviations due to dominance, σ_{Ce}^2 is the variance of environmental effects common to the household and σ_{Se}^2 is the variance of environmental effects special to the individual. The expected covariance between any two individuals i and j for the trait when this model is true is:

$$\text{cov}(i, j) = 2\phi_{ij}\sigma_a^2 + \Delta_{ij}\sigma_d^2 + R_{ij}\sigma_{Ce}^2 + I_{ij}\sigma_{Se}^2. \quad (2)$$

ϕ_{ij} , the kinship coefficient for individuals i and j , is the probability that an allele chosen at random from i is identical by descent with an allele at the same locus chosen at random from j . Δ_{ij} is Jacquard's (1974) Δ_7 , the probability that the two alleles of individual i are identical by descent to the two alleles of individual j at the same locus. R_{ij} is 1.0 if i and j reside in the same household and 0.0 if they do not. I_{ij} is set to 1.0 if $i = j$ and 0.0 if $i \neq j$.

The genetic component of the multifactorial determination of normal cholesterol variability is parameterized according to Fisher's (1918) polygenic model, which assumes Hardy-Weinberg equilibrium, linkage equilibrium, no epistasis, and no genotype-environment interaction. We assume that all individuals residing in a household share the same common environment and no inbreeding has occurred.

The likelihood of the model given the pedigree data is obtained by assuming a multivariate normal distribution for the trait among the members of the pedigree. The justification for this assumption is given by Lange (1978). Under this assumption, the likelihood (L) expressed as the natural log of the multivariate normal function, can be represented as:

$$L = -\frac{1}{2} \ln |\text{cov}| - \frac{1}{2} (X - \mu)' \text{cov}^{-1} (X - \mu) \quad (3)$$

(Rao, 1973). The elements of the cov matrix are functions of the four variance components given in equation 2 and the observed genealogical and household information. Cholesterol values for the members of a pedigree are represented by the vector X , while μ is the vector of their respective secretor type means. The means of the secretor blood type marker phenotypes (Secr. = dominant, Nonsecr. = recessive) are used in this study to distinguish two significantly different subpopulations of cholesterol values in Tecumseh (Sing & Orr, 1976). These differences have been found in three distinct populations (Longman *et al.* 1969; Beekman & Olivecrona, 1970; Sing & Orr, 1976) and suggest that this marker locus may be one of several markers which are either directly involved in cholesterol determination or closely linked to loci which modify cholesterol metabolism. Estimates of the two secretor means and four variance components were taken to be those values which maximize L .

The likelihood was maximized using Fisher's scoring algorithm. If $\gamma(t)$ is the current vector of the four variance components and two secretor means, Fisher's scoring algorithm replaces $\gamma(t)$ by:

$$\gamma(t+1) = \gamma(t) + I^{-1}S, \quad (4)$$

where the vector S contains the partial derivatives (scores) of L with respect to the parameters of the model. The information matrix (I) consists of the negatives of the second partial derivatives of L , with respect to these same parameters. These derivatives were first expressed in terms of the parameters of the multifactorial model by Lange *et al.* (1976). S and I are evaluated at $\gamma(t)$ for each pedigree and then are summed over all pedigrees for each iteration. The iterative procedure was terminated when the absolute difference between $\gamma(t)$ and $\gamma(t+1)$ was less than 10^{-5} for each parameter. A large number of initial vectors were considered to ensure that the procedure converged to the true maximum of the likelihood.

The sample

Total non-fasting serum cholesterol values for 9182 individuals were first ln transformed and then were adjusted for age, sex and socioeconomic status (Sing & Orr, 1977). The age adjustment was a third order polynomial equation. These concomitant factors account for approximately 38% of the variance of ln cholesterol among randomly sampled individuals from Tecumseh. A subsample of 3995 individuals on whom we had 12 blood markers and household information was used in this likelihood analysis. These individuals were distributed in 537 genetically independent pedigrees, ranging in size from 4 to 83 individuals. The distribution of cholesterol values in this subsample was representative of the distribution in the larger sample of 6366 individuals used in the analysis of correlations reported by Sing & Orr (1978).

To study the effect of sampling strategy on the estimated variance components, we selected five subsets of pedigrees to represent different distributions of pedigree size. Subset 1 comprised all pedigrees ranging in size from 10 to 15; a total of 568 individuals were included. Subset 2 was composed of 584 individuals, distributed among the 13 pedigrees of sizes 31–83. Subset 3 was taken to be the union of subsets 1 and 2. Subset 4 (referred to as the probability sample) contained 528 individuals in 58 pedigrees, ranging in size from 4 to 83 individuals, and was sampled to approximate the distribution of pedigree sizes in the total population. Subset 5 included the entire sample of 3995 individuals in the 537 pedigrees of size 4–83. We assume subset 5 represents the distribution of pedigree sizes in the Tecumseh population.

Table 1. *A description of the five subsets of pedigrees*

	Subset				
	1 (medium pedigrees)	2 (large pedigrees)	3 (subset 1 + subset 2)	4 (probability sample)	5 (all pedigrees)
Number of pedigrees ...	48	13	61	58	537
Size of pedigrees	10-15	31-83	10-15, 31-83	4-83	4-83
Number of individuals	568	584	1152	528	3995
In adjusted cholesterol					
Mean	5.2881	5.2916	5.2899	5.3054	5.2959
Variance	0.0268	0.0277	0.0272	0.0269	0.0278
Skewness (g_1)	-0.077	-0.337	-0.211	-0.539	-0.229
Secretor phenotype frequencies					
Secretor	0.7447	0.8168	0.7812	0.7727	0.7534
Nonsecretor	0.2553	0.1832	0.2188	0.2273	0.2466

The statistics

Standard descriptive statistics were computed on the non-fasting ln cholesterol values included in each of the subsets of pedigrees. The maximum likelihood estimates of σ_a^2 , σ_d^2 , σ_{Ce}^2 , σ_{Se}^2 , $\mu_{Secr.}$, and $\mu_{Nonsecr.}$, and their standard errors were computed for each subset of pedigrees. The standard errors of the estimates of the parameters were taken from the inverse of the information matrix, evaluated at the maximum likelihood estimates. The likelihood ratio test criterion was used to evaluate the hypothesis that one or more parameters were zero and the hypothesis that two subsets of pedigrees are described by the same set of parameter values. After multiplication by 2, the ratio of likelihoods is distributed approximately as a chi-square. To test the hypothesis that the estimates of the parameters obtained from each subsample were not significantly different from the estimates for the entire sample the ratio of the likelihood of each subsample, evaluated at its maximum point, to the likelihood of this same subsample evaluated at those estimates obtained from the total sample was computed. We assumed that this chi-square had 6 degrees of freedom.

RESULTS

A description of the subsets of pedigrees used in the analysis along with the cholesterol means and the frequencies of the secretor types are given in Table 1. Although the total distribution of ln cholesterol is not significantly skewed, the subsets studied had slight negative skewness. None of the subsets had a mean cholesterol value or secretor frequency which is substantially different from the other subsets or the total sample of pedigrees (subset 5).

The maximum likelihood estimates of variance components, their standard errors, heritability estimates, secretor type means and the ln likelihood values are given in Table 2 when the four variance components are estimated simultaneously (the complete model). The relative proportions of the various components provided from the least squares analysis are also in Table 2. Narrow and broad sense heritability estimates from the total sample of pedigrees agree closely

with estimates from the least squares analysis of the complete model (Sing & Orr, 1978). Pedigrees in subsets 1 and 2 do not yield significantly different estimates of the variance components ($\chi^2 = 1.02$, D.F. = 6). The vectors of parameter estimates from subsets 1, 2, 3 and 4 are not significantly different from subset 5 although the estimates of the additive and dominance variance components in subsets 1, 2 and 3 are higher than estimates based on the total sample of pedigrees. The narrow sense heritability estimated from the probability sample of pedigree sizes (subset 4) most closely resembles the estimate from the total sample of pedigrees. In every subset, the *t* test of the difference between secretor type means is highly significant.

Tests of hypotheses about variance components are given in Table 3 for the medium size pedigrees (subset 1), the probability sample (subset 4), and all pedigrees (subset 5). In all three subsets, we reject the hypothesis of no genetic influence on the variability of ln cholesterol ($\sigma_a^2 = \sigma_d^2 = 0$). We failed to reject any other hypotheses ($\sigma_d^2 = 0$; $\sigma_{C_e}^2 = \sigma_{S_e}^2 = 0$; $\sigma_{C_e}^2 = 0$; $\sigma_d^2 = \sigma_{C_e}^2 = 0$) for medium size pedigrees. Although not shown in Table 3, the inferences for pedigrees of size 31 to 83 individuals were identical to those obtained from the medium size pedigrees. For the probability sample we can reject all but the hypothesis of no dominance effects, while all hypotheses considered were rejected in the sample of pedigrees of sizes 4–83 individuals.

DISCUSSION

The maximum likelihood estimates of the variance components from all data in pedigrees of size 4–83 individuals are similar to the least squares estimates of Sing & Orr (1978), when the complete model with four variance components is considered. The large estimate of dominance variance (39.77%, $\chi^2 = 20.6$, 1 D.F., Table 3) estimated under the complete model, using the MLE procedure, is probably due to an incorrectly defined biological model and/or violations of the assumptions of the likelihood model on which the statistical analysis is based. Sing & Orr (1978) found that by redefining the common household environment to allow for heterogeneity of effects within and between generations the estimate of percent variance due to dominance effects was reduced to 7.9. Their redefined model took the form:

$$\sigma_p^2 = \sigma_a^2 + \sigma_d^2 + \sigma_{C_{epo}}^2 + \sigma_{S_{ep}}^2 + \sigma_{S_{ep}}^2 \quad (5)$$

for parents living together and:

$$\sigma_p^2 = \sigma_a^2 + \sigma_d^2 + \sigma_{C_{epo}}^2 + \sigma_{C_{ess}}^2 + \sigma_{S_{es}}^2 \quad (6)$$

for siblings living together. $\sigma_{C_{epo}}^2$ represents variance among households shared by parents and offspring. $\sigma_{C_{ep}}^2$ and $\sigma_{C_{ess}}^2$ are the variances among household effects shared by parents and sibs, respectively. $\sigma_{S_{ep}}^2$ and $\sigma_{S_{es}}^2$ are the variances of environmental factors special to parents and sibs, respectively. If all sibs reside in the same household, the estimates for σ_d^2 and $\sigma_{C_{ess}}^2$ would be confounded. However, in these data, some sibs reside in different households. When this reparametrization was considered in the least squares analysis, the estimate of $\sigma_{C_{epo}}^2$ became zero, while $\sigma_{C_{ess}}^2$ and $\sigma_{C_{ep}}^2$ were estimated to be 6.0 and 9.7% of the total phenotypic variance, respectively. The biological model employed in the likelihood procedure was redefined to conform more closely with the definition suggested by the least squares analysis. To do this the R_{ij} coefficients for parent offspring pairs were taken to be zero. The model becomes equation (1) with $\sigma_{C_{epo}}^2 = 0$, $\sigma_{C_e}^2 = \sigma_{C_{ep}}^2 = \sigma_{C_{ess}}^2$, and $\sigma_{S_e}^2 = \sigma_{S_{ep}}^2 = \sigma_{S_{es}}^2$. According to this redefinition of environment, the parents and offspring residing in the same house do not share common

Table 3. Hypothesis testing

	Hypothesis				
	No genetics ($\sigma_a^2 = \sigma_d^2 = 0$)	No dominance ($\sigma_d^2 = 0$)	No environment ($\sigma_{C_e}^2 = \sigma_{S_e}^2 = 0$)	No common environment ($\sigma_{C_e}^2 = 0$)	No dominance or common environment ($\sigma_d^2 = \sigma_{C_e}^2 = 0$)
Degrees of freedom for test ...	2	1	2	1	2
Subset 1, medium pedigrees, χ^2	24.62*	3.76	1.16	3.08	4.74
Subset 4, probability sample, χ^2	12.64*	0.20	10.54*	9.06*	9.20*
Subset 5, all pedigrees, χ^2	110.40*	20.60*	20.80*	20.00*	40.20*

* Statistically significant at the 0.01 level of probability.

Table 4. Variance components with parent and offspring common environment equal to 0.0

(A) All pedigrees size 4-83

	Sing & Orr (1978) Model H	Subset 5	Non-outlier pedigrees in subset 5	Outlier pedigrees in subset 5
Number of pedigrees ...		537	472	65
		Percentage		
σ_a^2	57.9	54.27	48.91	64.00
σ_d^2	7.9	9.42	1.71	23.22
$\sigma_{C_e}^2$	9.7*	10.25	8.96	11.53
$\sigma_{S_e}^2$	6.0†			
	24.5‡	26.06	40.42	1.25
	28.2§			

(B) Medium pedigrees size 10-15

	Subset 1	Non-outlier pedigrees in subset 1	Outlier pedigree in subset 1
Number of pedigrees ...	48	40	8
		Percentage	
σ_a^2	55.82	53.40	32.94
σ_d^2	39.47	18.50	67.06
$\sigma_{C_e}^2$	4.67	10.66	0.00
$\sigma_{S_e}^2$	0.04	17.44	0.00

* *Cepp*, † *Cess*, ‡ *Sep*, § *Ses*.

environmental effects. Any common household environmental effects are shared only by individuals within the same generation.

Another potential contribution to the large estimate of dominance variance obtained from the pedigree analysis may come from pedigrees with individuals whose cholesterol values are in the tails of the sample distribution. Such pedigrees could be segregating for a dominant allele with a megaphenic effect which contributed disproportionately to the estimate of σ_d^2 when the polygene model is falsely assumed to explain cholesterol variability. At least two such loci, with dominant alleles which determine a megaphenic effect, are thought to be segregating in the Tecumseh population (Sing *et al.* 1975). The estimate of σ_d^2 could also be inflated by the

Table 5. *Variance components estimated from outlier pedigrees with parent and offspring common environment equal to 0.0*

No. of individuals in pedigrees	21 pedigrees (outliers only in top 5% of sample distribution)	44 pedigrees (outliers in bottom 5% and outliers in both top and bottom 5% of sample distribution)
	139	704
		Percentage
σ_a^2	19.84	59.02
σ_d^2	80.12	22.67
$\sigma_{C^e}^2$	0.02	15.62
$\sigma_{S^e}^2$	0.02	2.69
ln likelihood	-165.46	-831.41

non-random distribution among pedigrees of environmental factors with major influences on individual values. This non-random distribution may or may not be associated with the failure of the data to be distributed according to the multivariate normal function. To study the effects of pedigrees with outliers, 65 of the 537 pedigrees (referred to as outlier pedigrees) were selected as a separate subsample because at least one individual in each pedigree had a cholesterol value in either the upper or lower 1% of the total sample. These individuals had an adjusted cholesterol value of less than 131 or greater than 293 mg/100 ml. Of these 65 pedigrees, 34 had at least one individual in the lower 1%, 29 had at least one individual in the top 1% and two pedigrees had individuals in both the upper and lower 1%. The individuals in these pedigrees represent 21% of the total sample of 3995 individuals.

The maximum likelihood estimates of the variance components, assuming that parents and offspring do not share common environments are given in Table 4.4 for the total 537 pedigrees, the 472 pedigrees without outliers, and the 65 outlier pedigrees. The estimate of 9.42% of variance due to dominance effects obtained using the 537 pedigrees agrees closely with the 7.9% obtained from Sing & Orr's (1978) best fitting model. In the 472 pedigrees without outliers, the percentage of σ_d^2 is 1.7%. Therefore, redefinition of the common environment reduced the contribution of dominance from 39.77 to 9.42%, while eliminating 65 outlier pedigrees reduced it further to 1.7%. An analysis of the outlier pedigrees will be discussed later.

Redefinition of the common environment changed the narrow sense heritability from 0.39 to 0.54, using all 537 pedigrees and to 0.49 using the 472 pedigrees with non-outliers. The estimate of 0.58, given by Sing & Orr (1978), may be slightly inflated due to inclusion of outliers. Other familial studies have reported heritabilities for cholesterol values in the range of 0.56–0.84 (Rao, 1976; Martin, Kurczynski & Steinberg, 1973; Christian & Kang, 1977; Heiberg, 1974).

The small but statistically significant dominance component (1.7%) in the sample of pedigrees without outliers could represent inter-locus non-additivity of polygenes or could be attributable to factors not specified in the model. Epistasis will inflate the dominance variance more than the additive variance when no epistatic term is included (Falconer, 1960). The linkage disequilibrium in this population (Sinnock & Sing, 1972) could also contribute to an upward bias

in the dominance variance (Robinson *et al.* 1960; Gallais, 1974; Weir & Cockerham, 1977). While we have no strong evidence to support the hypothesis that either assortative mating or selection is occurring in this population, the additive variance would be more affected than the dominance variance if these forces were operative (Falconer, 1960; Nei, 1960).

The comparison in Table 4A of the separate analysis of the 65 outlier pedigrees with the analysis of the remaining 472 pedigrees reveals that estimates of variance components are significantly different ($\chi^2 = 182.6$, D.F. = 6). Estimates of σ_a^2 , σ_d^2 , and σ_{Ce}^2 from the outlier pedigrees were all larger than those obtained from pedigrees without an outlier. The greater covariance observed between individuals in pedigrees with outliers could be explained by one of three possible hypotheses. First, cholesterol variability is determined by polygenes and household environment in all pedigrees, but an additional environmental factor is operating in the outlier pedigrees which causes an inflation in covariance between individuals only in these pedigrees. Secondly, the genetic variability of cholesterol in all pedigrees is explained mostly by polygenic factors, with the outliers identifying additional household factors which cause household variation to be greater in certain pedigrees. Thirdly, the genetic variance in the 472 pedigrees is explained by polygenes while the genetic variance in outlier pedigrees is controlled by one or more major loci with a polygenic background.

To understand the source of the increased covariance in the subset of 65 pedigrees, these pedigrees were first subdivided into three groups. A first group (high outliers) consisted of 139 individuals in 21 pedigrees with one or more individuals in each pedigree exceeding the 95th percentile of the total sample of 3995 individuals and no individuals in the lower 5%. A second group (low outliers) of 133 individuals in 21 pedigrees had one or more individuals below the 5th percentile of the total sample and none above the 95th percentile. A third group (mixed outliers) consisted of 571 individuals in 23 pedigrees with one or more individuals in both the top and bottom 5% of the sample. A nested analysis of variance (Sokal & Rohlf, 1969) among individuals within households, among households within pedigrees and among pedigrees was utilized to distinguish between the pedigree, household, or megaphenic explanations for the increase in covariance in these three groups of outlier pedigrees. Since there was no statistically significant variability among pedigree means, it is unlikely that the outliers are caused by a single major environmental factor which is common to an entire pedigree. The low outlier and mixed outlier groups each had statistically significant household effects. It is probable that the outliers in these pedigrees are members of households with significantly higher or lower mean values. On the other hand, no statistically significant household or pedigree effects were identified in the high outlier group. Using a modification of a test suggested by Fain & Ott (1976) designed to detect the presence of a locus with a major effect, we found a highly significant difference ($\alpha = 0.01$) between the within pedigree variance of these high outlier pedigrees and the variance among 2075 genetically independent individuals identified by Sing & Orr (1978). The within pedigree variance was 1.3 times higher than the variance among unrelated individuals. The results of the analysis of variance and the test for heterogeneity of variance suggest that a major locus is segregating only in the pedigrees with outliers having high cholesterol values.

Of the 44 pedigrees with individuals having cholesterol values in the top 5% of the sample (high outliers and mixed outliers), this analysis suggests 21 pedigrees (48%) are possibly segregating at one or more loci with a megaphenic effect. This finding is consistent with the

report by Goldstein *et al.* (1973) that a dominant allele with a major effect is segregating in 54 % of the pedigrees ascertained by survivors of myocardial infarctions, who also had cholesterol values which exceeded the 95th percentile. While the distribution of ln cholesterol was negatively skewed in the sample of 3995 individuals, evidence for a major gene was found only for the high cholesterol values. The estimates of variance components using the 21 pedigrees with outliers exceeding the 95th percentile are compared in Table 5 with those obtained from the remaining 44 outlier pedigrees. The large contribution of additive and dominance effects (99.97 %) and the relatively small effects of both the common and special environments in the group of 21 pedigrees also suggests that one or more loci with a major effect are the likely cause of outliers.

The comparison of variance estimates obtained from subsets of pedigrees of specific sizes with those obtained from all pedigrees (Tables 2 and 3) suggests that inferences about variance components may be markedly affected by a number of effects not accounted for by the biological model analysed. Neither medium (10–15 individuals) nor large (31–83 individuals) pedigrees of approximately 550 individuals gave reasonable estimates of variance components. Our analyses suggest that redefinition of the environmental parameters and removal of outlier pedigrees from the set of medium size pedigrees (Table 4B) resulted in estimates which were closer to those expected from the probability sample. These results suggest that the estimation procedure proposed by Lange *et al.* (1976) is sensitive to improper definition of the environmental model and genetic heterogeneity among pedigrees. On the other hand, if the environmental parameters are appropriately defined and pedigrees with outliers are eliminated, non-representative sampling of pedigree sizes did not lead to substantially different inferences about variance components. As can be seen from the work of Spence *et al.* (1978), random samples may show consistent estimates of narrow sense heritability while showing a wide range of broad sense heritabilities. While we are not advocating any particular sampling scheme, the choice of a random sample of pedigree sizes may more often avoid erroneous interpretation of the genetic and environmental influences on trait variability than the choice of a fixed pedigree size. The inferences from a random sample were closest to those from the total sample before the common environment was redefined or outlier pedigrees were identified.

CONCLUSIONS

The MLEs of variance components from pedigree data were similar to estimates obtained using a least squares analysis of the same data. Since a large sample was utilized (almost 4000 individuals) we cannot predict how similar the inferences from the two procedures might be on a smaller data set. A probability sample of pedigree sizes gave estimates closest to those obtained from the total sample of available pedigrees. Different samples of pedigree sizes can yield inferences similar to those obtained from an analysis of all available data only if the environmental model defines common environment separately for parents and offspring and pedigrees with outliers, suggestive of genetic and/or environmental heterogeneity, are removed.

This study confirms an earlier report that approximately 50 % of the variability in non-fasting cholesterol in Tecumseh is due to additive polygenic effects. In addition, the use of the MLE technique and pedigree data enabled us to study the heterogeneity of genetic and environmental sources of variability among subsets of the population. In doing so, 21 pedigrees were

identified whose cholesterol variability may be under the control of one or more loci with major effects on total cholesterol levels.

We are deeply indebted to Karlin Richardson for her help in assembling the pedigrees. We also thank J. L. Annest, T. Beaty, C. C. Cockerham, K. Lange, J. D. Orr, J. Ott, W. J. Schull, P. Smouse, M. A. Spence, and D. C. Rao for their helpful comments.

The work was supported by NIH Post-doctoral Fellowship 1F32HL05293, ERDA Contract E-(2828) and U.S. Public Health Service Program Project Grant H-6378 to the Cardiovascular Research Center, University of Michigan, Ann Arbor, Michigan.

REFERENCES

- BEEKMAN, L. & OLIVECRONA, T. (1970). Serum-cholesterol and ABO and Lewis blood groups. *Lancet* i, 1000.
- CHAKRABORTY, R., SCHULL, W. J., HARBURG, E., SCHORK, M. A. & ROEPER, R. (1977). Heredity, stress and blood pressure, a family set method. V. Heritability estimates. *J. Chron. Dis.* **30**, 683-699.
- CHRISTIAN, J. C. & KANG, K. W. (1977). Maternal influence on plasma cholesterol variation. *Am. J. Hum. Genet.* **29**, 462-467.
- ELSTON, R. C. (1975). On the correlation between correlations. *Biometrika* **62**, 133-148.
- ELSTON, R. C. & STEWART, J. (1971). A general model for the genetic analysis of pedigree data. *Hum. Hered.* **21**, 523-542.
- FAIN, P. R. & OTT, J. (1976). Heterogeneity of within sibship variance as a test of the major gene hypothesis (abstr.). In *Proceedings 5th International Congress of Human Genetics. Excerpta Medica* **397**, 180.
- FALCONER, D. C. (1960). *Introduction to Quantitative Genetics*. New York: Ronald Press.
- FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399-433.
- GALLAIS, A. (1974). Covariance between arbitrary relatives with linkage and epistasis in the case of linkage disequilibrium. *Biometrics* **30**, 429-446.
- GOLDSTEIN, J. L., HAZZARD, W. R., SCHROTT, H. G., BIERMAN, E. L. & MOTULSKY, A. G. (1973). Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. *J. Clin. Invest.* **52**, 1544-1568.
- HEIBERG, A. (1974). Heritability of serum lipids. In *Atherosclerosis III. Proceedings of the 3rd International Symposium* (ed. G. Schettler and A. Weizel). New York: Springer-Verlag.
- JACQUARD, A. (1974). *The Genetic Structure of Populations*. New York: Springer-Verlag.
- LANGE, K. (1978). Central limit theorems for pedigrees. (In preparation.)
- LANGE, K., WESTLAKE, J. & SPENCE, M. A. (1976). Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann. Hum. Genet., Lond.* **39**, 485-491.
- LANGMAN, M. J. S., ELWOOD, P. C., FOOTE, J. & PYRIE, D. R. (1969). ABO and Lewis blood-groups and serum-cholesterol. *Lancet* ii, 607-609.
- MARTIN, A. O., KURCZYNSKI, T. W. & STEINBERG, A. G. (1973). Familial studies of medical and anthropometric variables in a human isolate. *Am. J. Hum. Genet.* **25**, 581-593.
- MORRISON, D. F. (1965). *Multivariate Statistical Methods*. New York: McGraw-Hill.
- NANCE, W. & COREY, L. (1976). Genetic models for analysis of data from the families of identical twins. *Genetics* **83**, 811-826.
- NEI, M. (1963). Effect of selection on the components of genetic variance. In *Statistical Genetics and Plant Breeding* (ed. W. D. Hanson and H. F. Robinson), pp. 501-515. National Academy Science - National Research Council.
- OTT, J. (1978). In *Genetic Epidemiology* (ed. N. Morton and C. Chung). New York: Academic Press.
- RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.
- RAO, D. C. (1976). Path analysis of lipoproteins in nuclear families (abstr.). In *Proceedings 5th International Congress of Human Genetics. Excerpta Medica* **397**, 193.
- SCHULL, W. J., HARBURG, E., ERFURT, J. C., SCHORK, M. A. & RICE, R. (1970). A family set method for estimating heredity and stress. II. Preliminary results of genetic methodology in a pilot survey of Negro blood pressure, Detroit, 1966-67. *J. Chron. Dis.* **23**, 83-92.
- SING, C. F., CHAMBERLAIN, M. A., BLOCK, W. D. & FEILER, S. (1975). Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. I. Analysis of the frequency distribution for evidence of a genetic polymorphism. *Am. J. Hum. Genet.* **27**, 333-347.

- SING, C. F. & ORR, J. D. (1976). Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. III. Identification of genetic effects using 12 polymorphic genetic blood marker systems. *Am. J. Hum. Genet.* **28**, 453-464.
- SING, C. F. & ORR, J. D. (1978). Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. IV. Separation of polygene from common environment effects. *Am. J. Hum. Genet.* (In the Press.)
- SINNOCK, P. & SING, C. F. (1972). Analysis of multilocus genetic systems in Tecumseh, Michigan. II. Consideration of the correlation between non-alleles in gametes. *Am. J. Hum. Genet.* **24**, 393-415.
- SPENCE, M. A., WESTLAKE, J. & LANGE, K. (1977). Estimation of the variance components for dermal ridge count. *Ann. Hum. Genet., Lond.* **41**, 111-115.
- SOKAL, R. R., ROHLF, E. J. (1969). *Biometry: the Principles and Practice of Statistics in Biological Research*. San Francisco: W. H. Freeman.
- WEIR, B. & COCKERHAM, C. C. (1977). Two locus theory in quantitative genetics. In *Proceedings of the International Conference on Quantitative Genetics* (ed. E. Pollak, O. Kempthorne and T. E. Bailey Jr.). Iowa State University Press.