

# Univariate and Bivariate Analyses of Cholesterol and Triglyceride Levels in Pedigrees

Michael Boehnke, Patricia P. Moll, Kenneth Lange, William H. Weidman, and Bruce A. Kottke

*Departments of Biostatistics (M.B.), Epidemiology (P.P.M.), and Human Genetics (P.P.M), University of Michigan, Ann Arbor; Department of Biomathematics, University of California, Los Angeles (K.L.); Mayo Clinic and Mayo Foundation, Rochester, Minnesota (W.H.W., B.A.K.)*

A multivariate normal model for pedigree analysis is applied to fasting total serum cholesterol and total serum triglyceride measurements on 771 individuals in 95 pedigrees from Rochester, MN. Univariate and bivariate analyses are carried out to determine to what extent the aggregation and coaggregation in families of these two traits may be attributed to shared genetic and environmental factors. Pedigrees were ascertained through a sample of schoolchildren enriched for those with serum cholesterol levels in the highest and lowest deciles of their age- and sex-specific distributions. Ascertainment is corrected for by conditioning the likelihood on the trait values of the probands.

Univariate results confirm the findings of previous studies indicating that familial aggregation of serum cholesterol and triglyceride levels is due both to shared genes and to shared environmental factors. Results of the bivariate analyses suggest that the coaggregation of cholesterol and triglyceride levels in these families is strongly influenced by both shared genes (pleiotropy) and shared environmental factors. These findings are consistent with our understanding of lipid metabolism and of specific environmental factors known to influence both traits.

**Key words:** serum cholesterol, serum triglyceride, variance components, bivariate analysis, pedigrees, maximum likelihood

Received for publication January 15, 1985; revision received May 23, 1985.

Address reprint requests to Michael Boehnke, PhD, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109.

## INTRODUCTION

The degree to which shared genes and shared environmental factors influence the familial aggregation of fasting serum cholesterol and fasting serum triglyceride levels has received a great deal of attention in recent years because of the importance of these variables as measures of lipid metabolism and as risk factors in atherosclerosis [Robertson, 1981]. Family studies suggest that the familial aggregation of these traits is under the control of both genes and shared environmental factors, with genes relatively more important in determining cholesterol levels, and shared environmental factors more important for triglyceride levels [Sing and Orr, 1978; Rao et al, 1979; Iselius, 1979; Rao et al, 1982].

Cholesterol and triglyceride levels are significantly correlated within individuals in the general population in adults and children [Connor et al, 1982]. Here we ask to what degree can the correlation between the two traits be ascribed to the effects of common genes and common environmental factors? That is, do some of the genes and environmental factors that influence cholesterol levels also influence triglyceride levels? Studying the coaggregation of these traits in families provides the means to address these questions.

Other investigators have approached problems of bivariate genetic analysis by discriminant analysis [Namboodiri et al, 1975; Goldin et al, 1980], factor analysis [Martin and Eaves, 1977; Morton et al, 1978], and path analysis [Moll et al, 1978; Dal Colletto et al, 1981; Hanis, 1981; Darlu et al, 1982; Hanis et al, 1983; McGue et al, 1983; McGue, 1983; Vogler, 1985; Vogler and DeFries, 1985]. Here we analyze a sample of 3-generation pedigree data from Rochester, Minnesota. While in medical genetics pedigree usually refers to the diagram expressing the genetic relationships among individuals, here we use pedigree to describe a multigeneration family. We apply a multivariate normal form of the multifactorial model [Lange et al, 1976; Lange and Boehnke, 1983]. While this form of the model has seen wide application to univariate data [Spence et al, 1977; Moll et al, 1979; Annett et al, 1979; Beaty and Fajans, 1982; Rotter et al, 1982; Annett et al, 1983; Moll et al, 1983; Beaty et al, 1983], this paper presents its first application to bivariate data.

The multifactorial model assumes cholesterol and triglyceride levels are determined by the summed effects of many genes, each of small effect (polygenes), and a variety of environmental factors which may be shared by sets of pedigree members or be specific to each individual. Separate univariate pedigree analyses of cholesterol and triglyceride levels allow us to apportion the variability of each of the traits into components due to genes, environmental factors shared by pedigree members, and environmental factors specific to individuals. Subsequent bivariate pedigree analysis of cholesterol and triglyceride levels jointly allow us to apportion the covariability of the two traits into these various components.

## MATERIALS AND METHODS

### Sample

As part of a Pediatric Specialized Center of Research (SCOR) project, a cross-sectional survey of adult risk factors for atherosclerosis was carried out in 3,666 school children from Rochester, Minnesota, in 1973 and 1974. After the serum cholesterol data were collected on the school children, a 3% subsample was selected

on the basis of the deviations of the children from the mean for their age and sex. The sample was enriched for the upper tail of the distribution with approximately 40% randomly selected from the highest decile, 40% randomly selected from the middle 8 deciles, and 20% randomly selected from the lowest decile. Medical records were reviewed and school children with illness, those receiving medications that could secondarily affect serum lipid levels, and adoptees were excluded. The remaining eligible 100 white children ranging in age from 6 to 16 yr, were selected as probands. Two of these children were sibs; 1 of the 2 was randomly chosen to be excluded as a proband. Another proband had extreme hypertriglyceridemia (1,333 mg/dl) and was excluded. Three additional probands identified three separate pedigrees shown by complex segregation analysis to be segregating at a major locus for hypercholesterolemia [Moll et al, 1984]. When these three pedigrees were included in our analyses, a pedigree test statistic approach [Hopper and Mathews, 1982; Boehnke, 1983; Boehnke and Lange, 1984] suggested that their cholesterol data were not consistent with the multifactorial model, providing additional evidence for segregation at a major locus for cholesterol in those pedigrees. Since the multifactorial model does not allow for major locus effects, the pedigrees of these probands were excluded from subsequent analyses. Thus, the population of inference for this study excludes those rare individuals who exhibit high cholesterol values owing to segregation at a major locus for cholesterol.

The remaining 95 probands included 38 in the highest decile for total serum cholesterol for their age and sex, 20 in the lowest decile, and 37 in the middle 8 deciles. Fasting serum cholesterol and triglyceride measurements were made on the 95 probands and 676 of their biological relatives: 225 grandparents, 187 parents, and 264 full sibs. The number of pedigree members with measurements for both cholesterol and triglyceride ranged from 4 to 14, with a median of 8.

Measurements of plasma cholesterol and triglyceride levels were made after an overnight fast. Specific laboratory methods are described by Ellefson et al [1978]. Precision of the measurements was ensured by the participation of the Mayo Clinic SCOR Lipid Laboratory in the Center for Disease Control Cooperative Lipid Standardization Program.

## Model

Pedigree data were analyzed under the multifactorial model using the multivariate normal approach described by Lange et al [1976] and Lange and Boehnke [1983]. The model assumes that the trait(s) are the result of the summed effects of many genes, each of small effect (polygenes), together with the effects of a variety of environmental factors either shared by sets of pedigree members or specific to each individual. We further assume random mating, Hardy-Weinberg and linkage equilibrium, additivity both within and between loci (that is, no genetic dominance and no epistasis), and no genotype-environment interaction.

Following Lange et al [1976, 1983], we assume that a pair of quantitative traits (in our case, cholesterol and triglyceride levels) are measured on  $n$  members of a pedigree, giving pedigree trait vectors  $\underline{X} = (X_1, \dots, X_n)^t$  and  $\underline{Y} = (Y_1, \dots, Y_n)^t$ , respectively. Here, superscript  $t$  represents vector or matrix transpose. We assume that after adjustment for the effects of concomitants (such as age and sex), the pedigree trait vector ( $\underline{X}$  or  $\underline{Y}$  in the univariate case,  $\underline{Z} = (X_1, \dots, X_n, Y_1, \dots, Y_n)^t$  in the bivariate case) follows a multivariate normal distribution with mean vector zero and covariance

matrix described below. The covariance matrix depends on the relationships between the pedigree members and on the genetic and environmental factors included in the model. Theoretical justification for the assumption of multivariate normality has been given by Lange [Lange, 1978; Lange and Boehnke, 1983]. This assumption allows us to deal with complete pedigrees rather than only pairs of relatives as in the alternative approach of path analysis [Wright, 1921; Li, 1975].

Here, we assume that variability in cholesterol and triglyceride levels is due to the summed effects of additive polygenes, environmental factors shared by members of the same household, of the same sibship, and of the same spouse pair, and an environmental factor specific to each individual which may also reflect measurement error. Under these assumptions, the  $n \times n$  covariance matrix for cholesterol is

$$\Omega_{xx} = \sigma_{axx} 2\Phi + \sigma_{sxx}S + \sigma_{hxx}H + \sigma_{pxx}P + \sigma_{cxx}I.$$

$\sigma_{axx}$  is the additive genetic variance,  $\sigma_{sxx}$  the shared sibship environmental variance,  $\sigma_{hxx}$  the shared household environmental variance,  $\sigma_{pxx}$  the shared spouse environmental variance, and  $\sigma_{cxx}$  the individual-specific environmental variance for cholesterol.  $2\Phi$ , S, H, P, and I are the  $n \times n$  known covariance matrices representing the theoretical correlations between the pedigree members for each of the factors considered. Under this scheme, the cholesterol total variance for any pedigree member is  $\sigma_{axx} + \sigma_{sxx} + \sigma_{hxx} + \sigma_{pxx} + \sigma_{cxx}$ . Analogous expressions hold for the covariance matrix and the within-individual total variance for triglyceride ( $Y$ ), with the subscript  $yy$  replacing the subscript  $xx$ .

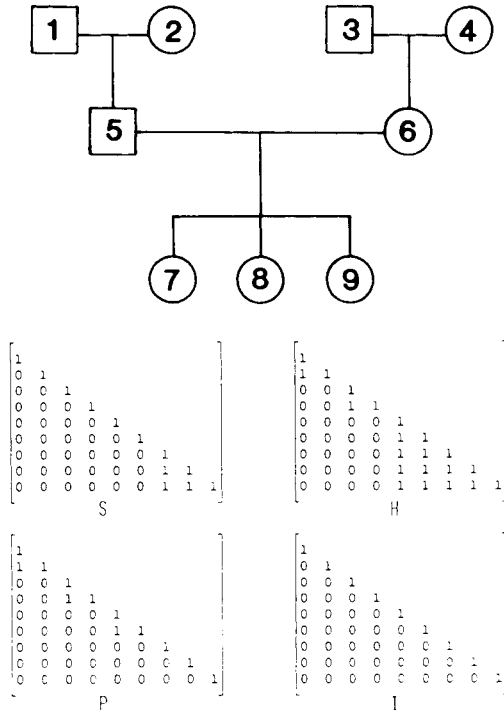


Fig. 1. Covariance matrices for the given pedigree corresponding to shared sibship (S), shared household (H), shared spouse (P), and individual-specific (I) environments. Note: since these matrices are symmetric, only the lower triangles are given.

For the additive genetic variance, the covariance matrix is  $2\Phi$ , twice the kinship matrix. The kinship matrix has  $ij$ th entry equal to the kinship coefficient of individuals  $i$  and  $j$  (that is, the probability that genes drawn at random from the same locus from  $i$  and  $j$  are identical by descent [Malecot, 1969; Jacquard, 1974]).

For each environmental factor, we partition the pedigree into disjoint and exhaustive blocks [Lange and Boehnke, 1983]. Given a partition of the pedigree, we assume that all individuals within the same block are exposed to the same environment, and that environmental contributions to the several blocks are independent and identically distributed. The corresponding covariance matrices have  $ij$ th element equal to 1 if individuals  $i$  and  $j$  are in the same block, and 0 if they are in different blocks. Figure 1 illustrates the partitioning of a representative 3-generation pedigree into blocks appropriate for the environmental factors considered in our application to cholesterol and triglyceride levels by defining the matrices  $S$ ,  $H$ ,  $P$ , and  $I$ . Note that the partitions corresponding to  $S$  and  $P$  both represent subpartitions of the partition corresponding to  $H$ . Since all the sibs in each pedigree in this study were living together at the time of data collection, shared sibship environmental effects could not be separated in our analyses from genetic dominance. Therefore, when we conclude below that shared sibship environmental effects are important for these traits, the possibility that the effect is at least partly due to genetic dominance cannot be ruled out. Similarly, in a cross-sectional study with no information on duration of cohabitation or marriage, the possible effects of shared spouse environmental factors can not be distinguished from assortative mating. More subtle environmental parameterizations, taking into account the cohabitation history of pedigree members, are possible [Hopper and Mathews, 1982; Hopper and Mathews, 1983; Hopper and Culross, 1983; Lange, 1985].

In the bivariate case, we again consider variance components and corresponding covariance matrices for cholesterol and triglyceride separately. In addition, we now model the between-trait cross covariances. That is, we allow not only for the possibility that the cholesterol and triglyceride measurements for two pedigree members may be correlated, but also for the possibility that the cholesterol measurement of one pedigree member may be correlated with the triglyceride measurement of the other.

For bivariate data, it is convenient to describe the covariance matrix  $\Omega$  of the trait vector  $\underline{Z} = (X_1, \dots, X_n, Y_1, \dots, Y_n)^t$  using partitioned matrices [Rao, 1973]. With assumptions as before,

$$\Omega = \begin{bmatrix} \Omega_{xx} & \Omega_{xy} \\ \Omega_{yx} & \Omega_{yy} \end{bmatrix}$$

$\Omega_{xx}$  and  $\Omega_{yy}$  are  $n \times n$  matrices of the same form as the univariate covariance matrices for cholesterol and triglyceride, respectively.  $\Omega_{xy} = \Omega_{yx}$  are identical  $n \times n$  symmetric matrices of the same linear form as  $\Omega_{xx}$  and  $\Omega_{yy}$ , except that here the known covariance matrices are multiplied by covariance components rather than variance components. That is,

$$\Omega_{xy} = \Omega_{yx} = \sigma_{axy}2\Phi + \sigma_{sxy}S + \sigma_{hxy}H + \sigma_{pxy}P + \sigma_{exy}I,$$

where  $\sigma_{rxy}$  is an unknown covariance component ( $r = a, s, h, p, e$ ). The choice of  $\Omega_{xy} = \Omega_{yx}$  corresponds to the assumption that the expected covariance between the

cholesterol value  $X_i$  for individual  $i$  and the triglyceride value  $Y_j$  for individual  $j$  should equal the expected covariance between  $X_j$  and  $Y_i$ . Again, the within-individual covariance between the two traits is just  $\sigma_{axy} + \sigma_{sxy} + \sigma_{hxy} + \sigma_{pxy} + \sigma_{exy}$ .

A difficulty with using cross covariances is that the natural bounds on these parameters (namely, that they be in magnitude no greater than the square root of the product of the corresponding variance components) are of a complicated non-linear form. Therefore, we have chosen to reparameterize the model in terms of cross correlations, which must by definition lie between -1 and 1. Since the method of maximum likelihood is invariant to reparameterization [Cox and Hinkley, 1974], this choice does not affect our inferences. Parameterizing in terms of cross correlations is carried out by replacing each covariance component  $\sigma_{rxy}$  by  $\rho_{rxy} \sigma_{rxx}^{1/2} \sigma_{ryy}^{1/2}$ , where  $\rho_{rxy}$  is the corresponding cross correlation component. Details of this reparameterization are given in Appendix 1.

Under the assumptions given, the natural logarithm of the pedigree likelihood for bivariate data may be written

$$L = -\frac{1}{2} \log |\Omega| - \frac{1}{2} \underline{Z}' \Omega^{-1} \underline{Z}, \quad (1)$$

where  $|\Omega|$  is the determinant of  $\Omega$ . For univariate data, we replace  $\underline{Z}$  and  $\Omega$  in the above equation with  $\underline{X}$  and  $\Omega_{xx}$  or  $\underline{Y}$  and  $\Omega_{yy}$ . For a sample of independent pedigrees, the loglikelihood of the sample is the sum of the pedigree loglikelihoods. Maximizing the loglikelihood permits maximum likelihood parameter estimation and testing of hypotheses under the likelihood ratio criterion.

Loglikelihoods were maximized using a variable metric maximization routine written by Kenneth Lange and based on a paper by Powell [1978]. Convergence of the maximization process was tested using a convergence criterion first suggested by Hopper and Mathews [1982], and generalized by us in Appendix 2. Approximate standard errors (SE) were obtained by inversion of minus the matrix of numerically-calculated second partial derivatives evaluated at the maximum likelihood parameter estimates.

### Ascertainment Correction

When pedigrees are ascertained through probands, it is important to correct for the sampling procedure. To this end, Hopper and Mathews [1982] suggested conditioning on the trait values of the probands [Cannings and Thompson, 1977] for this multivariate normal form of the multifactorial model. Since ascertainment of our pedigrees was nearly single, with only one pedigree ascertained through as many as two probands, we chose to condition on the probands' trait values as means of ascertainment correction. Conditioning on the trait value of the probands requires only that we replace the (unconditional) multivariate normal loglikelihood (1) with the conditional multivariate normal loglikelihood of the non-proband trait values given the proband trait value [Hopper and Mathews, 1982; Boehnke, 1983; Appendix 1].

## RESULTS

The mean total cholesterol and triglyceride levels among the 676 relatives of the 95 probands were 205.5 mg/dl (SD = 51.1 mg/dl) and 107.6 mg/dl (SD = 69.7 mg/dl), respectively. The within-individual correlation between cholesterol and tri-

glyceride values was .475 in the relatives, .275 in probands, and .476 overall. Prior to fitting the multifactorial model, the cholesterol and triglyceride measurements on the relatives and probands were transformed to univariate normality by an inverse normal scores transformation [Daniel and Wood, 1971]. While transformation to univariate normality in no way guarantees multivariate normality for the pedigree trait vectors, it is a reasonable step in that direction. Transformation to multivariate normality is not feasible. We normalized the data prior to regression on concomitants (see below) because of the significant skewness and kurtosis in the raw data.

Normalized lipid measurements for the relatives were regressed on the best-fitting cubic polynomial of age separately for men, women, and women using oral contraceptives. Reduction in variance due to regression was 36%, 46%, and 41% for men, women, and women using oral contraceptives, respectively for cholesterol, and 35%, 33%, and 5% in the same groups for triglycerides. Normalized proband values were then adjusted using the same regression equations. The correlation between transformed values was .370 in the relatives, .300 in the probands, and .355 overall. After normalization and regression, trait means and variances were not significantly heterogeneous between sexes or across 10-yr age strata. The within-individual trait correlation between transformed cholesterol and transformed triglyceride was not significantly heterogeneous between sexes; it was marginally heterogeneous across age strata ( $\chi^2 = 14.30$ ,  $df = 7$ ,  $p = .05$ , [Rao, 1973]), but showed no particular trend with age. Skewness and kurtosis for transformed cholesterol and triglyceride were not significant at the .05 level, whether or not the probands were included.

Results of the univariate analyses for the transformed cholesterol and triglyceride values are presented in Tables I and II. For cholesterol, comparing parameter estimates with their SE under the complete model suggested the presence of significant effects owing to additive polygenes, shared sibship environment, and individual-specific environment, and the possibility of an effect of a household shared environment. The shared spouse component bounded at zero. Excluding the additive polygenic effect and considering a purely environmental model resulted in a  $\chi^2$  statistic of

TABLE I. Cholesterol Univariate Analyses

Model	Model Parameters*					2(L <sub>c</sub> -L) <sup>a</sup>
	$\sigma_{axx}$	$\sigma_{sxx}$	$\sigma_{hxx}$	$\sigma_{pxx}$	$\sigma_{exx}$	
Complete	.202 (.049)	.054 (.025)	.050 (.027)	.000 <sup>b</sup>	.215 (.034)	0.00
$\sigma_{sxx}, \sigma_{hxx}, \sigma_{pxx}, \sigma_{exx}$	—	.070 (.030)	.128 (.026)	.000 <sup>b</sup>	.325 (.026)	18.98
$\sigma_{axx}, \sigma_{sxx}, \sigma_{exx}$	.255 (.042)	.071 (.026)	—	—	.195 (.031)	3.90
$\sigma_{axx}, \sigma_{exx}$	.264 (.037)	—	—	—	.248 (.026)	13.26
$\sigma_{sxx}, \sigma_{exx}$	—	.218 (.033)	—	—	.320 (.025)	56.03
$\sigma_{exx}$	—	—	—	—	.545 (.030)	135.89

\*Maximum likelihood parameter estimates (approximate SE).

<sup>a</sup>Likelihood ratio statistic for comparison with complete model. L<sub>c</sub>, L are the loglikelihoods for the complete and reduced models, respectively.

<sup>b</sup>Parameter bounded at zero.

TABLE II. Triglyceride Univariate Analyses

Model	Model Parameters*					2(L <sub>c</sub> -L) <sup>†</sup>
	$\sigma_{\text{a}}_{\text{yy}}$	$\sigma_{\text{s}}_{\text{yy}}$	$\sigma_{\text{h}}_{\text{yy}}$	$\sigma_{\text{p}}_{\text{yy}}$	$\sigma_{\text{c}}_{\text{yy}}$	
Complete	.133 (.073)	.066 (.034)	.020 (.042)	.052 (.060)	.330 (.089)	0.00
$\sigma_{\text{s}}_{\text{yy}} \cdot \sigma_{\text{h}}_{\text{yy}} \cdot \sigma_{\text{p}}_{\text{yy}} \cdot \sigma_{\text{c}}_{\text{yy}}$	—	.068 (.033)	.083 (.026)	.000 <sup>b</sup>	.450 (.032)	3.47
$\sigma_{\text{a}}_{\text{yy}} \cdot \sigma_{\text{s}}_{\text{yy}} \cdot \sigma_{\text{c}}_{\text{yy}}$	.148 (.044)	.077 (.031)	—	—	.376 (.037)	2.64
$\sigma_{\text{a}}_{\text{yy}} \cdot \sigma_{\text{c}}_{\text{yy}}$	.192 (.038)	—	—	—	.410 (.035)	10.42
$\sigma_{\text{s}}_{\text{yy}} \cdot \sigma_{\text{c}}_{\text{yy}}$	—	.154 (.031)	—	—	.449 (.032)	15.88
$\sigma_{\text{c}}_{\text{yy}}$	—	—	—	—	.616 (.033)	59.13

\*Maximum likelihood parameter estimates (approximate SE).

<sup>†</sup>Likelihood ratio statistic for comparison with complete model. L<sub>c</sub>, L are the loglikelihoods for the complete and reduced models, respectively.

<sup>b</sup>Parameter bounded at zero.

18.98 on 1 degree of freedom ( $p < .0005$ ), confirming the importance of the additive genetic effect. Considering the reduced model with shared household effect excluded, the reduction in loglikelihood suggested that the small effect owing to a shared household environment was also statistically significant ( $\chi^2 = 3.90$ ,  $df = 1$ ,  $.04 < p < .05$ ). Further reduction of the model to exclude either the polygenic effect or the shared sibship environmental effect resulted in a significant decrease in the loglikelihood. Thus, the most parsimonious model for explaining the observed variability in transformed cholesterol in these pedigrees included an additive polygenic effect and shared sibship, shared household, and individual-specific environmental effects. For this model, the estimated polygenic heritability for transformed cholesterol was  $.202 / (.202 + .054 + .050 + .215) = .388$ .

For transformed triglyceride (Table II), the individual-specific environmental effect was clearly significant. However, evidence for the other effects was less strong for transformed triglyceride than for transformed cholesterol. Under the complete model, the additive genetic and shared sibship environmental terms appeared most important, both in absolute terms and in comparison with their approximate SE. Excluding the additive polygenic effect and considering a purely environmental model resulted in a  $\chi^2$  statistic of 3.47, suggesting an important, though not quite statistically significant genetic effect. The fact that the shared spouse environmental term bounded at zero complicates the assessment of statistical significance in this case. Exclusion of the remaining two factors, shared household and shared spouse environments, did not significantly reduce the loglikelihood ( $\chi^2 = 2.64$ ,  $df = 2$ ,  $p > .20$ ). After excluding these factors, further reduction of the model to exclude either the additive genetic effect or the shared sibship environmental effect resulted in a significant decrease in the loglikelihood. We therefore chose a parsimonious model for transformed triglyceride including an additive polygenic effect and shared sibship and individual-specific environmental effects. For this model, the estimated polygenic heritability for transformed triglyceride was  $.148 / (.148 + .077 + .376) = .246$ .

We began the bivariate analysis with a complete model including variance components corresponding to the factors identified as significant in the univariate



TABLE III. Cholesterol and Triglyceride Bivariate Analyses

Model	Model Parameters*										2(L <sub>c</sub> -L) <sup>a</sup>
	$\sigma_{axx}$	$\sigma_{sxx}$	$\sigma_{hxx}$	$\sigma_{exx}$	$\sigma_{ayy}$	$\sigma_{syy}$	$\sigma_{eyy}$	$\rho_{axy}$	$\rho_{sxy}$	$\rho_{exy}$	
$\sigma_{axx}, \sigma_{sxx}, \sigma_{hxx}, \sigma_{exx},$ $\sigma_{ayy}, \sigma_{syy}, \sigma_{eyy},$ $\rho_{axy}, \rho_{sxy}, \rho_{exy}$	.209 (.049)	.060 (.026)	.036 (.023)	.213 (.033)	.155 (.045)	.079 (.032)	.368 (.037)	.487 (.141)	.699 (.216)	.250 (.079)	0.00
$\sigma_{axx}, \sigma_{sxx}, \sigma_{exx},$ $\sigma_{ayy}, \sigma_{syy}, \sigma_{eyy},$ $\rho_{axy}, \rho_{sxy}, \rho_{exy}$	.252 (.042)	.072 (.026)	—	.196 (.031)	.155 (.045)	.079 (.032)	.368 (.037)	.443 (.127)	.654 (.201)	.262 (.082)	2.74

\*Maximum likelihood parameter estimates (SE)

<sup>a</sup>Likelihood ratio statistic for comparison with complete model. L<sub>c</sub>, L are the loglikelihoods for the complete and reduced models, respectively.

analyses, together with cross correlations corresponding to those factors identified as significant in both univariate analyses. That is, we assumed that factors which did not influence the familial aggregation of both the traits separately could not influence their coaggregation.

Variance component estimates and their approximate SE in the bivariate analyses (Table III) were very similar to the corresponding values in the univariate analyses (Tables I and II). The only notable difference was that the shared household environmental term for cholesterol was somewhat reduced. Estimates for the additive polygenic, shared sibship environmental, and individual-specific environmental cross correlations were all significantly greater than zero, whether or not the marginally significant shared household effect for cholesterol was included (Table III). Each estimated cross correlation was at least 3 SE greater than 0. Eliminating any of the 3 cross correlations resulted in a significant decrease in the loglikelihood (data not shown). For the additive cross correlation,  $\chi^2 = 8.11$ ,  $p < .005$ . For the sibship environmental cross correlation,  $\chi^2 = 5.64$ ,  $p < .025$ . For the individual-specific environmental cross correlation,  $\chi^2 = 7.39$ ,  $p < .01$ . Equivalent significance levels for the cross correlations were obtained when the shared household environmental term for cholesterol was excluded.

While cross correlation component estimates were affected to a certain extent by the inclusion of the shared household environmental variance for cholesterol (Table III), the partitions of the between-trait covariance which they implied were not. Noting that the *r*th covariance component may be calculated as  $\sigma_{rxy} = \rho_{rxy}\sigma_{rxx}^{1/2}\sigma_{ryy}^{1/2}$ , we found that the estimated cross covariances were nearly unchanged by the presence or absence of the shared household term for cholesterol. With the term included,  $\hat{\sigma}_{axy} = .088$ ,  $\hat{\sigma}_{sxy} = .048$ , and  $\hat{\sigma}_{exy} = .070$ . With the shared household term excluded,  $\hat{\sigma}_{axy} = .088$ ,  $\hat{\sigma}_{sxy} = .049$ , and  $\hat{\sigma}_{exy} = .070$ . In either case, the pleiotropic effects of genes common to both traits were estimated to account for 42–43% of the within-individual covariance for transformed cholesterol and triglyceride, while shared sibship and individual-specific environmental effects common to both traits were estimated to account for 23–24% and 34%, respectively.

## DISCUSSION

### Lipid Metabolism and Environmental Factors

Our findings that common genetic factors are important in the coaggregation of serum cholesterol and serum triglyceride levels are consistent with knowledge of the

metabolism of these plasma lipids. Chylomicrons and very low density lipoprotein (VLDL) particles carry most of the triglyceride transported in plasma and moderate amounts of cholesterol. Most low density lipoprotein (LDL) particles are end products of VLDL catabolism and carry approximately 70% of plasma cholesterol in normal individuals. Direct secretion of LDL by the liver may also occur. It is not known how much cholesterol enters the circulation on chylomicrons and VLDL and how much cholesterol is attached to LDL in the plasma through the actions of the plasma enzyme lecithin cholesterol acetyltransferase (LCAT) and the cholesterol exchange protein.

Given the complexity of lipid metabolism, it is reasonable to hypothesize that many genes with small effects are involved, and that at least some of these genes affect both cholesterol and triglyceride levels. Population studies have reported associations between small differences in cholesterol levels and certain polymorphic loci, notably ABO and secretor [Sing and Orr, 1976]. Individuals with blood group A or AB individuals appear to have higher rates of type II hyperlipidemia (elevated LDL), while individuals with blood group B or O have higher rates of type IV hyperlipidemia (elevated VLDL) [Morton, 1976].

Similarly, a number of environmental factors, including cigarette smoking and dietary and exercise habits, are known to affect one or both of these serum lipid levels in the adult population. Variability in carbohydrate, alcohol, and overall caloric intake, and in body weight are known to be associated with variability in triglyceride levels [Hartung et al, 1980]. Reduced intake of saturated fats and increased intake of polyunsaturated fats are associated with lower levels of both cholesterol and triglyceride [Ernst and Levy, 1980]. It appears that exercise levels can exert a strong effect on triglyceride levels, and a lesser effect on cholesterol levels [Hartung et al, 1980]. In addition, cholesterol and particularly triglyceride levels are subject to a certain amount of within-individual variability [Jacobs and Barrett-Connor, 1982]. These and other environmental factors may be contributing to the individual-specific and shared environmental effects noted in the univariate and bivariate analyses.

### Ascertainment

Cannings and Thompson [1977] demonstrated that conditioning on probands provides an appropriate ascertainment correction when ascertainment is single. Hopper and Mathews [1982] pointed out that this ascertainment correction could be applied to the multivariate normal form of the multifactorial model. While conditioning on probands has been applied in segregation analysis, its application to the multifactorial model to date had been limited to simulated data [Boehnke, 1983; Boehnke and Lange, 1984]. Typically, data analysis under the multifactorial model has assumed random sampling even if the data were non-randomly ascertained [Rao et al, 1979; Beaty and Fajans, 1982]. Ascertainment correction in path analysis has only recently been considered [Hanis and Chakraborty, 1984].

Previously, we carried out a simulation study in which pedigree data were generated and analyzed under the multivariate normal form of the multifactorial model after ascertainment through the upper 5% of the distribution [Boehnke, 1983; Boehnke and Lange, 1984]. In that simulation, analysis of ascertained pedigrees by conditioning on probands consistently resulted in underestimates of the additive genetic variance, with corresponding increase in the individual-specific environmental variance, even when ascertainment was single. These biases may have arisen owing to the difficulty of finding a suitable transformation of the ascertained data to approximate multivariate normality [Hanis and Chakraborty, 1984]. We were con-

cerned that the results of the current analyses might be similarly biased. Several facts argue that this was not the case.

First, ascertainment of the current data set was not through a single tail of the distribution. While ascertainment did result in a sample enriched for probands from the highest and lowest age- and sex-specific deciles, probands came from the entire range of the distribution. Second, the results of our univariate analyses were quite comparable to the results of other studies in which pairs of relatives or families were randomly sampled [Sing and Orr, 1978; Iselius, 1979; Rao et al, 1982].

Third, our results were only slightly affected when two alternative approaches to dealing with ascertainment were employed. Our alternatives were to either 1) exclude the probands from the data and, after normalization and regression, analyze the pedigrees as if randomly sampled, or 2) ignore ascertainment, include the probands, and again, after normalization and regression, analyze the pedigrees as if randomly sampled. For cholesterol, excluding the probands generally resulted in slightly smaller estimates of the additive genetic effect and the shared sibship and shared household environmental effects, and slightly larger values for the individual-specific and shared spouse environmental effects, in comparison with conditioning on probands. Ignoring ascertainment had the reverse effect. The only difference in inference was that without the probands, the shared household effect for cholesterol dropped below the level of statistical significance. For triglyceride, differences were negligible. In the bivariate analyses, ignoring ascertainment gave equivalent results to conditioning on probands. Excluding the probands removed the evidence for a significant shared sibship cross correlation. This might reflect an important difference in the two data sets since the probands were not representative of the pedigree as a whole; a more likely alternative is that it simply reflects the few sib pairs present in the data when the 95 probands, of the total of  $264 + 95 = 359$  sibs in 95 pedigrees, were excluded.

### Previous Studies

Our approach to bivariate modeling assumes that the covariability of cholesterol and triglyceride is due to common genetic and environmental factors. This approach corresponds to the path analytic approaches of Dal Colletto et al [1981] and McGue [1983]. Dal Colletto et al [1981] analyzed cholesterol, triglyceride, and lipoprotein fraction measurements on 105 healthy twin pairs in Brazil. They reported a significant genetic cross correlation for cholesterol and triglyceride. However, they failed to allow for the possibility of an individual-specific environmental cross correlation. It seems probable *a priori* that such an effect would exist; by excluding this effect, the genetic effect was very likely overestimated. In addition, Dal Colletto et al estimated that heritability was .700 for cholesterol and .803 for triglyceride. Both these values, particularly that for triglyceride, are uncommonly high. These facts make their results difficult to interpret.

McGue [1983] analyzed cholesterol and triglyceride data in 160 randomly sampled nuclear families, some ascertained through juvenile, others through adult offspring. He found that a common environmental effect was required to explain the coaggregation of cholesterol and triglyceride levels in the families, and that either a genetic or an individual-specific environmental cross correlation was required as well; however, he could not distinguish between the two. While his parameterization of common environmental effects is not directly comparable to ours, his results appear similar, though less definitive. Several reasons could account for this differ-

ence in resolution. First, our sample was somewhat larger, 95 probands and 676 relatives as compared to roughly 530 relatives in his study. Second, our 3-generation pedigrees will have provided more contrasts in the data than given by his nuclear families. Third, since the method of maximum likelihood is statistically efficient [Cox and Hinkley, 1974], employing a maximum likelihood approach on complete pedigrees will have provided greater information for the analysis than could be provided by summarizing data by correlations and then analyzing those correlations as in path analysis. Empirical support for this theoretical statement is provided by Moll and Sing [1979] and Annett et al [1979].

An alternative path analytic approach to bivariate modeling was described by Hanis [Hanis, 1981; Hanis et al, 1983] and applied to weight and blood pressure. Their approach assumes that one trait directly affects the other at the phenotypic level. Their regression approach, while in some ways more general than our model, requires longitudinal data; even with longitudinal data, not all effects may be jointly estimable [Hanis, 1981; Hanis et al, 1983].

## CONCLUSION

The results of our bivariate analyses suggest that the covariability of cholesterol and triglyceride levels within individuals may be ascribed to (1) the pleiotropic effects of common genes, (2) the common effects of environmental factors shared by sibs, and (3) the common effects of environmental factors specific to the individual. Thus, while human geneticists tend to think of pleiotropy in terms of major loci, here is an example of polygenes with pleiotropic effects on cholesterol and triglyceride levels. We conclude that the factors, genetic and environmental, that determine cholesterol variability are not independent of the genetic and environmental factors that influence triglyceride variability in our sample of pedigrees.

In the current study, we have analyzed data on 771 Caucasians from Rochester, MN. The recent work of the Collaborative Lipid Research Clinics [1984] provides an ideal opportunity to repeat this sort of analysis on a much larger data set representing many areas of North America. It would be interesting to determine whether results similar to those of the current study would be obtained in this much larger and more representative population.

## ACKNOWLEDGMENTS

We thank Dr. Charles Sing for helpful suggestions and criticisms in the course of this work, and Drs. Ralph Ellefson and Hymie Gordon from the Mayo Clinic for their contributions to this study. We also thank Mrs. Patrice Somerville for expert manuscript preparation. This research was supported in part by grants HL-24489 and AM-33329 and Research Career Development Award KO4-HD00307 from the National Institute of Child Health and Human Development and by Department of Energy contract DEAC0282ER60089.

## REFERENCES

- Annett JL, Sing CF, Biron P, Mongeau JG (1979): Familial aggregation of blood pressure and weight in adoptive families. II. Estimation of the relative contributions of genetic and common environ-

- mental factors to blood pressure correlations between family members. *Am J Epidemiol* 110:492–503.
- Annest JL, Sing CF, Biron P, Mongeau JG (1983): Familial aggregation of blood pressure and weight in adoptive families. III. Analysis of the roles of shared genes and shared household environment in explaining family resemblance for height, weight, and selected weight/height indices. *Am J Epidemiol* 117:492–506.
- Beaty TH, Fajans SS (1982): Estimating genetic and non-genetic components of variance for fasting glucose levels in pedigrees ascertained through non-insulin dependent diabetes. *Ann Hum Genet* 46:355–362.
- Beaty TH, Self SG, Chase GA, Kwiterovich PO (1983): Assessment of variance components models on pedigrees using cholesterol, low-density, and high-density lipoprotein measurements. *Am J Med Genet* 16:117–129.
- Boehnke M (1983): Advances in pedigree analysis: Ascertainment, goodness of fit, and optimization. PhD dissertation, University of California, Los Angeles.
- Boehnke M, Lange K (1984): Ascertainment and goodness of fit of variance component models for pedigree data. In Rao DC, Elston RC, Kuller LH, Fienleib M, Carter C, Havlik R (eds): "Genetic Epidemiology of Coronary Heart Disease." New York: Alan R. Liss, Inc., pp 173–192.
- Cannings C, Thompson EA (1977): Ascertainment in the sequential sampling of pedigrees. *Clin Genet* 12:208–212.
- Connor SL, Connor WE, Sexton G, Calvin L, Bacon S (1982): The effects of age, body weight and family relationships on plasma lipoproteins and lipids in men, women, and children of randomly selected families. *Circulation* 65:1290–1298.
- Cox DR, Hinkley DV (1974): "Theoretical Statistics." New York: Halstead Press, pp 287, 304.
- Dal Colletto GMD, Krieger H, Magalhaes JR (1981): Estimates of genetical and environmental determinants of serum lipid and lipoprotein concentrations in Brazilian twins. *Hum Hered* 31:232–237.
- Daniel C, Wood FS (1971): "Fitting Equations to Data." New York: John Wiley and Sons, pp 28–29.
- Darlu P, Rao DC, Henrotte JG, Lalouel JM (1982): Genetic regulation of plasma and red blood cell magnesium concentrations in man. I. Univariate and bivariate path analyses. *Am J Hum Genet* 34:874–887.
- Ellefsen RD, Elveback LR, Hodgson PA, Weidman WH (1978): Cholesterol and triglycerides in serum lipoproteins of young persons in Rochester, Minnesota. *Mayo Clin Proc* 53:307–320.
- Ernst N, Levy RI (1980): Diet, hyperlipidemia and atherosclerosis. In Goodhart RS, Shils ME (eds): "Modern Nutrition in Health and Disease." Philadelphia: Lea and Febiger, pp 1045–1070.
- Family Study Committees for the Lipid Research Clinics Program (1984): The collaborative Lipid Research Clinics program family study I. Study design and description of data. *Am J Epidemiol* 119:931–943.
- Goldin LR, Elston RC, Graham JB, Miller CH (1980): Genetic analysis of von Willebrand's disease in two large pedigrees: A multivariate approach. *Am J Med Genet* 7:279–293.
- Hanis CL (1981): Multivariate models for human genetic analyses: Development and application to systolic blood pressure and weight. PhD dissertation, University of Michigan, Ann Arbor.
- Hanis CL, Chakraborty R (1984): Non-random sampling in human genetics: Familial correlations. *IMA J Math Appl Med Biol* 1:193–213.
- Hanis CL, Sing CF, Clarke WR, Schrott HG (1983): Multivariate models for human genetic analysis: Aggregation, coaggregation, and tracking of systolic blood pressure and weight. *Am J Hum Genet* 35:1196–1210.
- Hartung GH, Foreyt JP, Mitchell RE, Vlasek I, Gotto AM (1980): Relation of diet to high-density-lipoprotein cholesterol in middle-aged marathon runners, joggers, and inactive men. *N Engl J Med* 302:357–361.
- Hopper JL, Culross PR (1983): Covariation between family members as a function of cohabitation history. *Behav Genet* 13:459–471.
- Hopper JL, Mathews JD (1982): Extensions to multivariate normal models for pedigree analysis. *Ann Hum Genet* 46:373–383.
- 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.
- Iselius L (1979): Analysis of family resemblance for lipids and lipoproteins. *Clin Genet* 15:300–306.
- Jacobs DR, Barrett-Connor E (1982): Retest reliability of plasma cholesterol and triglyceride: The Lipid Research Clinics prevalence study. *Am J Epidemiol* 116:878–885.

- Jacquard A (1974): "The Genetic Structure of Populations." New York: Springer-Verlag, p 8.
- Lange K (1978): Central limit theorems for pedigrees. *J Math Biol* 6:59-66.
- Lange K (1985): Cohabitation, convergence, and environmental covariance. Submitted to *Theo Pop Biol*.
- Lange K, Boehnke M (1983): Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. *Am J Med Genet* 14: 513-524.
- Lange K, Westlake J, Spence MA (1976): Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann Hum Genet* 39:485-491.
- Li CC (1975): "Path Analysis-a Primer." Pacific Grove, California: Boxwood Press.
- Malecot G (1969): "The Mathematics of Heredity." San Francisco: W.H. Freeman, p 8.
- Martin NG, Eaves LJ (1977): The genetical analysis of covariance structure. *Heredity* 38:79-95.
- McGue M (1983): Bivariate path analysis of plasma lipids. *Hum Hered* 33:145-152.
- McGue M, Rao DC, Reich T, Laskarzewski P, Gleuck CJ, Russell JM (1983): The Cincinnati lipid research clinic family study. Bivariate path analyses of lipoprotein concentrations. *Genet Res Camb* 42:117-135.
- 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 hypercholesterolemia. *Am J Hum Genet* 36:197-211.
- Moll PP, Harburg E, Burns TL, Schork MA, Ozgoren F (1983): Heredity, stress and blood pressure, a family set approach: The Detroit project revisited. *J Chron Dis* 36:317-328.
- Moll PP, Powsner R, Sing CF (1979): Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. V. Variance components estimated from pedigrees. *Ann Hum Genet* 42:343-354.
- Moll PP, Sing CF (1979): Sampling strategies for the analysis of quantitative traits. In Sing CF, Skolnick M (eds): "Genetic Analysis of Common Diseases." New York: Alan R. Liss, pp 307-342.
- Moll PP, Sing CF, Brewer GJ, Gilroy TE (1978): Multivariate analysis of the genetic effects on red blood cell glycolysis. In Brewer G (ed): "The Red Cell." New York: Alan R. Liss, pp 385-405.
- Morton NE (1976): Genetic markers in atherosclerosis: A review. *J Med Genet* 13:81-95.
- Morton NE, Gulbrandsen CL, Rhoads GG, Kagan A, Lew R (1978): Major loci for lipoprotein concentrations. *Am J Hum Genet* 30:583-589.
- Namboodiri KK, Elston RC, Glueck CJ, Fallat R, Buncher CR, Tsang R (1975): Bivariate analyses of cholesterol and triglyceride levels in families in which probands have type IIb lipoprotein phenotype. *Am J Hum Genet* 27:454-471.
- Powell MJD (1978): A fast algorithm for nonlinearly constrained optimization calculations. In Dold A, Eckmann B (eds): "Numerical Analysis." Berlin: Springer-Verlag, pp 144-157.
- Rao CR (1973): "Linear Statistical Inference and its Applications." New York: John Wiley, pp 17, 32, 432-436, 522.
- Rao DC, Laskarzewski PM, Morrison JA, Khoury P, Kelly K, Wette R, Russell J, Glueck CJ (1982): The Cincinnati research clinic family study: Cultural and biological determinants of lipids and lipoprotein concentrations. *Am J Hum Genet* 34:888-903.
- Rao DC, Morton NE, Gulbrandsen CL, Rhoads GG, Kagen A, Yee S (1979): Cultural and biological determinants of lipoprotein concentrations. *Ann Hum Genet* 42:467-477.
- Robertson FW (1981): The genetic component in coronary heart disease: A review. *Genet Res Camb* 37:1-16.
- Rotter JI, Wong FL, Samloff IM, Varis K, Siurala M, Ihama T, Ellis A, McConnell RB (1982): Evidence for a major dominance component in the variation of serum pepsinogen I levels. *Am J Hum Genet* 34:395-401.
- Sing CF, Orr JD (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 CF, Orr JD (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* 30:491-504.
- Spence MA, Westlake J, Lange K (1977): Estimation of the variance components for dermal ridge count. *Ann Hum Genet* 41:111-115.
- Vogler GP (1985): Multivariate path analysis of familial resemblance. *Genet Epidemiol* 2:35-53.
- Vogler GP, DeFries JC (1985): Bivariate path analysis of familial resemblance for reading ability and

symbol processing speed. *Behav Genet* 15:111-121.  
 Wright S (1921): Correlation and causation. Part I. Method of path coefficients. *J Agr Res* 20:557-585.

Edited by John M. Opitz and James F. Reynolds

**APPENDIX 1: PARAMETERIZATION USING CROSS CORRELATIONS**

Parameterizing the multifactorial model in terms of cross correlations makes parameter constraints easier to deal with. The price is that the expression for the covariance matrix  $\Omega$  is no longer a linear function of the parameters. In addition, if the loglikelihood is to be maximized, modified expressions for the elements of the score vector (the first partial derivatives of the loglikelihood with respect to the parameters) are required.

In the covariance-component parameterization, we may write

$$\Omega = \sum_r \sigma_{rxx} \Omega_{rxx} + \sum_r \sigma_{ryy} \Omega_{ryy} + \sum_r \sigma_{rxy} \Omega_{rxy}, \tag{A1}$$

where  $\sigma_{rxx}$  is the *r*th cholesterol variance component,  $\sigma_{ryy}$  is the *r*th triglyceride variance component, and  $\sigma_{rxy}$  is the *r*th between-trait cross covariance. In this equation, the covariance matrices  $\Omega_{rxx}$ ,  $\Omega_{ryy}$ , and  $\Omega_{rxy}$  are  $2n \times 2n$  symmetric block matrices of the forms

$$\begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \quad \begin{bmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{B} & \mathbf{0} \end{bmatrix},$$

respectively, where  $\mathbf{0}$  represents the  $n \times n$  matrix of zeros and  $\mathbf{B}$  is the  $n \times n$  matrix of expected covariances appropriate to the *r*th parameter type ( $2\Phi$  for the additive genetic variance,  $\mathbf{S}$  for the shared sibship environmental variance, and so forth). Reparameterizing in terms of cross correlations requires that we replace  $\sigma_{rxy}$  in equation (A1) by  $\rho_{rxy} (\sigma_{rxx}\sigma_{ryy})^{1/2}$ .

As pointed out by Lange and Boehnke [1983], elements of the score vector are of the form

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2} \text{tr} (\Omega^{-1} \frac{\partial \Omega}{\partial \theta}) + \frac{1}{2} \underline{Z}' \Omega^{-1} \frac{\partial \Omega}{\partial \theta} \Omega^{-1} \underline{Z} \tag{A2}$$

where  $\theta$  is any covariance parameter.

Under the parameterization (A1),  $\partial \Omega / \partial \sigma_{rij} = \Omega_{rij}$ . Switching to the alternative parameterization using cross correlations,

$$\frac{\partial \Omega}{\partial \theta_{rii}} = \Omega_{rii} + \frac{1}{2} \rho_{rxy} (\sigma_{rij} / \sigma_{rii})^{1/2} \Omega_{rxy} \quad ij = xy \text{ or } yx$$

$$\frac{\partial \Omega}{\partial \rho_{rxy}} = (\sigma_{rxx}\sigma_{ryy})^{1/2} \Omega_{rxy}.$$

The obvious linearity of the trace function and quadratic forms in (A2) imply that the

score elements for the cross-correlation parameterization may be written as

$$\begin{aligned}\frac{\partial L}{\partial \sigma_{rii}} &= S_{rii} + \frac{1}{2} \rho_{rxy} (\sigma_{rij}/\sigma_{rii})^{1/2} S_{rxy} \quad ij = xy \text{ or } yx \\ \frac{\partial L}{\partial \rho_{rxy}} &= (\sigma_{rxx}\sigma_{ryy})^{1/2} S_{rxy},\end{aligned}\tag{A3}$$

where the  $S$ 's represent the score elements under the covariance-component parameterization.

## APPENDIX 2: AN EXTENDED COVERAGE CRITERION FOR MULTIVARIATE NORMAL MODELS

Hopper and Mathews [1982] demonstrated that for univariate traits and randomly-sampled pedigrees, a necessary condition for convergence for the multivariate normal form of the multifactorial model is

$$\Sigma (\underline{X} - \underline{\mu})^t \Omega^{-1} (\underline{X} - \underline{\mu}) = \Sigma n,$$

where  $\underline{\mu}$  and  $\Omega$  are the estimated mean vector and covariance matrix of  $\underline{X}$  at the maximum likelihood estimates, and  $n$  is the number of observations in  $\underline{X}$ . Both sums in the above equation run over all pedigrees in the data, and the pedigree subscript has been suppressed.

An analogous formula holds for bivariate (or more generally,  $q$ -variate) traits on randomly-sampled pedigrees and on ascertained pedigrees if ascertainment is corrected for by conditioning on probands [Cannings and Thompson, 1977]. If  $q$  is the number of traits observed on each pedigree member ( $q = 2$  in the bivariate case) and  $p$  is the number of probands in a pedigree (with  $p = 0$  if the pedigree was randomly sampled), then at the maximum

$$\Sigma (\underline{Z}_2 - \underline{\mu}_{2,1})^t \Omega_{22,1}^{-1} (\underline{Z}_2 - \underline{\mu}_{2,1}) = q \Sigma (n - p).\tag{A4}$$

Here,  $\underline{\mu}_{2,1}$  and  $\Omega_{22,1}$  are the estimated conditional mean vector and conditional covariance matrix of the trait vector  $\underline{Z}_2$  for the non-probands conditional on the trait vector  $\underline{Z}_1$  of the probands. That is, the sum of these pedigree quadratic forms evaluated at the maximum likelihood estimates over all pedigrees must equal the total number of trait values observed on the non-probands in the data. We will prove this result for the covariance-component parameterization. Without loss of generality we assume the trait vector  $\underline{Z}$  has mean vector  $\underline{0}$ . It should be noted that the proof given by Hopper and Mathews [1982] does not carry over directly to the case of ascertained pedigrees. Their proof requires the covariance matrix be a linear function of the variance components; the conditional covariance matrix  $\Omega_{22,1}$  is not.

In the absence of probands the proof of (A4) proceeds just as in Hopper and Mathews [1982]. It simplifies notation to lump all pedigrees into one super pedigree of independent subpedigrees. This causes  $\Omega$  to be block diagonal with nonzero blocks corresponding to the subpedigrees. Letting  $\theta_r$  denote a generic variance component, the linear relation (A1) can be recast as



$$\Omega = \sum_r \theta_r \frac{\partial}{\partial \theta_r} \Omega. \tag{A5}$$

At the maximum likelihood estimates the score vanishes, and so

$$\begin{aligned} 0 &= \sum_r \theta_r \frac{\partial}{\partial \theta_r} L \\ &= -\frac{1}{2} \text{tr} \left\{ \Omega^{-1} \sum_r \theta_r \left( \frac{\partial}{\partial \theta_r} \Omega \right) \right\} + \frac{1}{2} \underline{Z}' \Omega^{-1} \sum_r \theta_r \left( \frac{\partial}{\partial \theta_r} \Omega \right) \Omega^{-1} \underline{Z} \\ &= -\frac{1}{2} \text{tr}(\Omega^{-1} \Omega) + \frac{1}{2} \underline{Z}' \Omega^{-1} \underline{Z} \\ &= -\frac{1}{2} \text{tr}(\mathbf{I}) + \frac{1}{2} \underline{Z}' \Omega^{-1} \underline{Z} \\ &= -\frac{1}{2} mq + \frac{1}{2} \underline{Z}' \Omega^{-1} \underline{Z}, \end{aligned}$$

where  $\mathbf{I}$  is an identity matrix of size  $mq \times mq$ ,  $m = \sum n$  being the total number of people in the sample. An obvious algebraic rearrangement yields (A4) provided we take into account that  $\Omega^{-1}$  is block diagonal and that we have assumed the mean vector  $\underline{\mu}$  is zero.

When there are probands in the sample, by a suitable permutation of people and traits,  $\underline{Z}$  and  $\Omega$  can be brought into the partitioned forms

$$\underline{Z} = \begin{bmatrix} \underline{Z}_1 \\ \underline{Z}_2 \end{bmatrix}$$

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$$

where the subscripts 1 and 2 refer to the proband and non-proband parts of the sample respectively. The conditional loglikelihood may be written

$$L^* = L - L_1, \tag{A6}$$

where  $L_1$  is the loglikelihood of the probands. Since (A5) must hold for the whole sample as well as the proband sample, one has at the maximum likelihood estimates

$$\begin{aligned} 0 &= \sum_r \theta_r \frac{\partial}{\partial \theta_r} L^* \\ &= \sum_r \theta_r \frac{\partial}{\partial \theta_r} L - \sum_r \theta_r \frac{\partial}{\partial \theta_r} L_1 \\ &= -\frac{1}{2} mq + \frac{1}{2} \underline{Z}' \Omega^{-1} \underline{Z} + \frac{1}{2} m_1 q - \frac{1}{2} \underline{Z}'_1 \Omega^{-1}_1 \underline{Z}_1, \end{aligned}$$

where  $m$  is number of people in the whole sample and  $m_1 = \sum p$  is the number of probands.

To complete the proof we must show

$$\underline{Z}^t \Omega^{-1} \underline{Z} - \underline{Z}_1^t \Omega_{11}^{-1} \underline{Z}_1 = (\underline{Z}_2 - \mu_{2,1})^t \Omega_{22,1}^{-1} (\underline{Z}_2 - \mu_{2,1}). \quad (\text{A7})$$

The conditional mean vector  $\mu_{2,1}$  and the conditional covariance matrix  $\Omega_{22,1}$  given by Rao [1973] are

$$\begin{aligned} \mu_{2,1} &= \Omega_{21} \Omega_{11}^{-1} \underline{Z}_1 \\ \Omega_{22,1} &= \Omega_{22} - \Omega_{21} \Omega_{11}^{-1} \Omega_{12}. \end{aligned}$$

The proof of (A4) now follows upon taking logarithms in the determinant identity [Rao, 1973]

$$|\Omega| = |\Omega_{22,1}| \cdot |\Omega_{11}|$$

and using the expansion (1) of the text, which is valid for all three loglikelihoods in (A6). Note that the determinant terms cancel in the three loglikelihoods of (A6), leaving precisely the identity (A7).