*Genet. Res. Camb.* (1983), **42**, *pp.* 117–135
*With 1 text-figure*
*Printed in Great Britain*



# The Cincinnati Lipid Research Clinic family study

## Bivariate path analyses of lipoprotein concentrations

BY M. MCGUE,* D. C. RAO,* T. REICH,† P. LASKARZEWSKI,‡
C. J. GLUECK,‡ AND J. M. RUSSELL*

(*Received 23 September 1982 and in revised form 16 December 1982*)

SUMMARY

Methods for the analysis of the joint transmission of two phenotypes are described and used to determine the extent to which lipoprotein concentrations share a common genetic and/or environmental background. Analysis of data on 160 Caucasian nuclear families revealed that the observed phenotypic association between high-density cholesterol (HDL) and low-density cholesterol (LDL) could be accounted for in terms of common family environmental effects alone (estimated genetic correlation, $\rho_G = -0{\cdot}132 \pm 0{\cdot}136$; estimated residual environmental correlation, $\rho_R = 0{\cdot}065 \pm 0{\cdot}230$). The association between HDL and very-low-density cholesterol (VLDL) could not be accounted for in terms of family environmental effects alone. For HDL and VLDL the residual environmental correlation was significant while the genetic correlation was not ($\rho_G = -0{\cdot}111 \pm 0{\cdot}214$, $\rho_R = -0{\cdot}421 \pm 0{\cdot}172$). The correlation between LDL and VLDL also could not be accounted for in terms of common family environmental effects alone, although here a genetic relationship appears to be the important factor ($\rho_G = 0{\cdot}330 \pm 0{\cdot}192$, $\rho_R = 0{\cdot}010 \pm 0{\cdot}217$).

## 1. INTRODUCTION

In a recent analysis of within-family associations of five lipids and lipoproteins (total cholesterol; total triglyceride; high-density cholesterol, HDL; low-density cholesterol, LDL; and very-low-density cholesterol, VLDL) in 160 Caucasian families, part of the Cincinnati Princeton School District Family Study, it was concluded that: 'Whatever the relative contributions of genetics and environment are to the association of lipid-lipoprotein values between parents and offspring and between siblings, the magnitude of the association (except for high-density lipoprotein cholesterol) appears to outlast the period of shared household environment' (Morrison *et al.* 1982). In a subsequent analysis of the observed family resemblance for the lipid and lipoprotein variables with particular regard to the

* Division of Biostatistics, Washington University Medical School Box 8067, 4566 Scott Avenue, St Louis, MO 63110.

† Washington University Department of Psychiatry and The Jewish Hospital of St Louis, 216 South Kingshighway, St Louis, MO 63108

‡ The Lipid Research Clinic, The General Clinical Research Center, The CLINFO Center, University of Cincinnati, College of Medicine, 234 Goodman Street C2-3, Cincinnati, Ohio 45267.

resolution of cultural and biological inheritance it was further concluded that:

'On the whole, every trait gives highly significant evidence for both genetic and cultural inheritance with the single exception that triglyceride fails to support genetic effects. The non-transmitted sibship environment ($B$) has significant effects on all traits. Combined effect of assortative mating and cohabitation ($u$) is marginally significant for all traits except total cholesterol. Neither intergenerational differences in heritabilities nor maternal effects are significant for any trait, even though the latter exist (estimates of $f_M$ are consistently larger than those of $f_F$). Indices turn out to be good estimates of the indexed environment (estimates of $i$ and $iv$ are often close to 1·0). . . we took $y = z = 1$, $u = 0$ and $f_F = f_M$ (no intergenerational differences, no assortative mating and no maternal effects) as the most parsimonious hypothesis for each trait' (Rao *et al.* 1982*a*).

Given that the three lipoprotein variables, all of which are known risk factors for atherosclerotic disease, are considerably associated, it becomes important to resolve not only their individual but also their joint cultural and biological inheritance. Although multivariate formulations are possible, there are at least two reasons to prefer simpler bivariate methods. First, given an observed phenotypic association, the most relevant questions are (*a*) Is that association a result of a single common genotype? and (*b*) Is that association only a result of common environmental factors? (i.e. uncorrelated genetic factors). As will be demonstrated, bivariate path models allow for the testing of these hypotheses. Secondly, multivariate models are necessarily complex, often involving parameters and specifications not central to the investigation at hand. Consequently, meaningful data analysis becomes much more difficult in a multivariate system.

In the present paper a bivariate path model, an extension of the model considered by Darlu *et al.* (1982), and an elaboration of the model used for the analysis of twin data by Colletto, Krieger & Magalhaes (1981) is presented and used as the basis for the analysis of lipoprotein family data from the Cincinnati Family Study. Other researchers have developed alternative models of multivariate phenotypic transmission (e.g. Hanis, 1981; Hanis & Sing, 1981; Lange, Boehnke & Spence, 1983; Plomin & Defries, 1979; Reeve, 1952; Eaves & Gale, 1974): methods which will be briefly described and distinguished from the present formulation in a later section.

## 2. THE POPULATION STUDIED

The Cincinnati Lipid Research Clinic (LRC) Princeton School District Family Study (1976–8) (Morrison *et al.* 1982*a*) was a part of the National Heart, Lung and Blood Institute's multicenter collaborative program designed to assess the familial aggregation of lipids and lipoprotein levels (Heiss *et al.* 1980). Briefly, the Princeton School District Population Study was an epidemiological survey of lipids, lipoproteins and other coronary heart-disease risk factors in a biracial population of school children in grades 1–12 and their parents. Following the first two visits of the prevalence study (Morrison *et al.* 1978), a subgroup of probands was drawn from Visit 1 of this larger prevalence population for the Family Study. All first-degree relatives and spouses of selected probands were contacted; socio-

demographic data, fasting plasma lipids and lipoproteins, and clinical chemistry measurements were obtained. Probands for the Family Study included both randomly selected subjects and hyperlipidemic (top decile cholesterol or triglyceride) subjects (Kelley *et al.* 1983; Morrison *et al.* 1982*a*, *b*; Tyroler *et al.* 1979).

Here data on 160 Caucasian families ascertained through the randomly selected probands is analysed. There were a few three-generation families which were split into the component nuclear families, avoiding duplications whenever possible. Relatives of the probands studied include spouses, children, and sibs. Very few adopted relatives and half-sibs were studied, and due to very small sample sizes, they are not analysed here. More details of the population studied can be found in Laskarzewski *et al.* (1983) and Morrison *et al.* (1982*a*, *b*).

## 3. TRANSFORMATION OF DATA

Age and sex have substantial effects on plasma lipoprotein concentrations, while behavioural, social and physiological factors such as obesity, hematocrit, special diet, smoking and alcohol consumption generally have somewhat smaller effects (Laskarzewski *et al.* 1983; Gulbrandsen *et al.* 1977). The Cincinnati LRC family data was first adjusted for age and sex effects by stepwise multiple regression (Laskarzewski *et al.* 1983). Specifically, each lipoprotein variable was separately regressed on sex, age, age$^2$, age$^3$, sex × age, sex × age$^2$, sex × age$^3$, contraceptive usage, obesity as measured by Quetelet index (weight/height$^2$), hematocrit and special diet usage in a stepwise fashion retaining only the significant terms (Laskarzewski *et al.* 1983):

$$X = f(A, S, C) + g(Z) + e,$$

where $X$ = lipoprotein variable, $f(A, S, C)$ = polynomial involving age, sex and contraceptive terms, $g(Z)$ = linear function of obesity, hematocrit and special diet usage and $e$ = residual error.

A cubic function of the type considered here sufficed to eliminate age and sex effects almost completely. After fitting the above equation, the age–sex–contraceptive adjusted lipoprotein variable ($P$) and an index of familial environment ($I$) were defined as

$$P = X - \hat{f}(A, S, C)$$

and

$$I = \hat{g}(Z),$$

where $\hat{f}$ and $\hat{g}$ denote the estimated contributions. The part of the environment so estimated by the index ($I$) is called *family environment* or *indexed environment*, and the remainder is the *residual environment*. In this way two variables, $P$ and $I$, were generated for each of the three lipoprotein concentrations.

## 4. THE MODEL

The general bivariate path model is presented in Fig. 1. The six variables of the model, both observable and unobservable, are defined in Table 1, and the 18 basic parameters are defined in Table 2. As in the univariate models, it is assumed that the genotype, residual environment and family environment act additively to produce a phenotype. All interactions are assumed to be negligible.

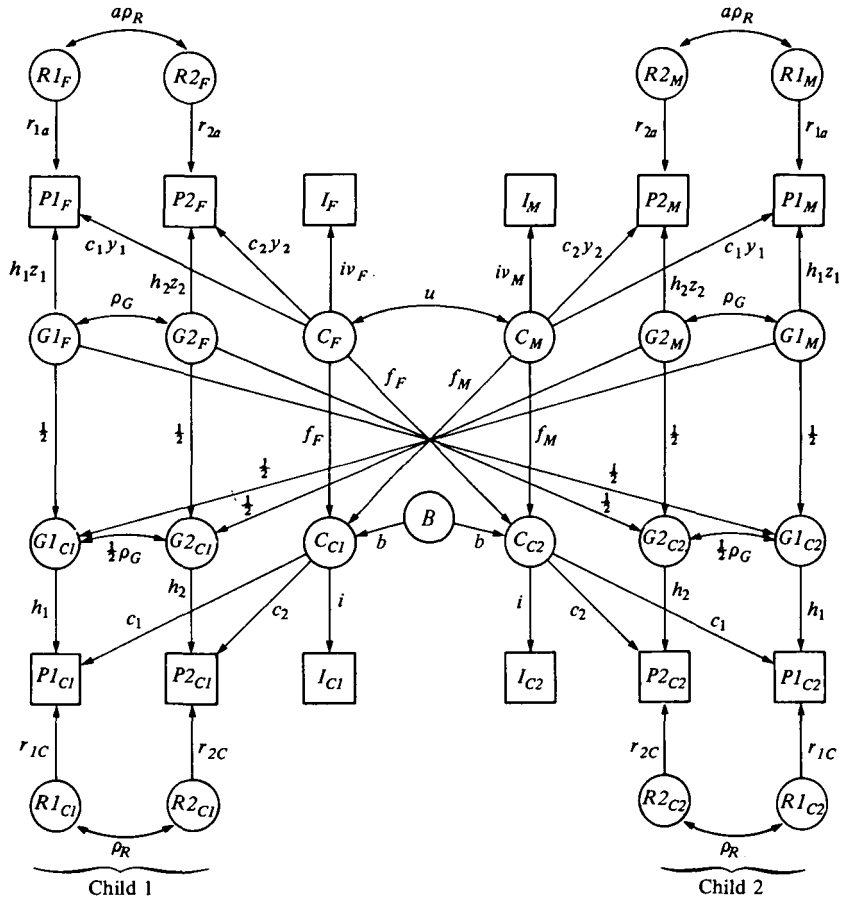


Fig. 1. Bivariate path model showing cultural and biological inheritance for two correlated phenotypes *P1* and *P2*, with corresponding genotypes *G1* and *G2*, a common familial environment *C*, and residuals *R1* and *R2*, *B* denotes non-transmitted common sibship environment, and the index *I* is an estimate of *C*. Subscripts *F*, *M*, *C1* and *C2* denote father, mother, and two children respectively.

Table 1. *Definition of variables in Fig. 1*

| Variable* | Definition |
|---|---|
| Observable | |
| *P1*, *P2* | Two correlated phenotypes |
| *I* | Index of family environment common to both phenotypes |
| Non-observable | |
| *G1*, *G2* | Correlated genotypes of *P1*, *P2* |
| *C* | Family environment common to both phenotypes |
| *B* | Non-transmitted common sibship environment |
| *R1*, *R2* | Non-transmitted residual environmental components of *P1* and *P2* |

* *F*, *M*, *C1*, *C2* denote father, mother, and two children respectively.

https://doi.org/10.1017/S0016672300021595 Published online by Cambridge University Press

Table 2. *Definition of the 18 basic and 4 derived parameters of the bivariate path model (Fig. 1)*

| Parameter | Definition |
|---|---|
| $h_i$ | Effect of $i$th genotype on the $i$th phenotype of a child (square root of genetic heritability of the $i$th trait); $i = 1, 2$ |
| $h_i z_i$ | Effect of $i$th genotype on the $i$th phenotype of an adult; $i = 1, 2$ |
| $c_i$ | Effect of environment on the $i$th phenotype of a child (square root of cultural heritability of the $i$th trait); $i = 1, 2$ |
| $c_i y_i$ | Effect of environment on the $i$th phenotype of an adult; $i = 1, 2$ |
| $u$ | Correlation between parental environments |
| $b$ | Effect of non-transmitted common sibship environment on child's environment |
| $f_F$ | Effect of father's environment on that of a child he rears |
| $f_M$ | Effect of mother's environment on that of a child she rears |
| $i$ | Effect of environment on child's index |
| $iv_F$ | Effect of environment on father's index |
| $iv_M$ | Effect of environment on mother's index |
| $\rho_G$ | Correlation between the two genotypes |
| $\rho_R$ | Correlation between the two non-transmissible residuals of a child |
| $a\rho_R$ | Correlation between the two non-transmissible residuals of an adult |
| Derived parameters | |
| $r_{ia}$ | Effect of $i$th residual upon $i$th phenotype of an adult; $r_{ia} = (1-h_i^2 z_i^2 - c_i^2 y_i^2)^{\frac{1}{2}}$ $(i = 1, 2)$ |
| $r_{ic}$ | Effect of $i$th residual upon $i$th phenotype of a child; $r_{ic} = (1-h_i^2 - c_i^2)^{\frac{1}{2}}$ $(i = 1, 2)$ |

The genetic factors for the two phenotypes are delineated in terms of two separate genotypes which are correlated ($\rho_G$). Environmental factors are represented in terms of non-transmissible residual environmental components, one for each trait (R1 and R2), which are correlated ($\rho_R$ in children and $a\rho_R$ in adults), and a single transmissible family environment ($C$) common to both phenotypes. A delineation of family environmental factors similiar to that used for genetic and residual environmental factors, although possible, may not be of much value as the same index variables are often used to estimate the family environment of the two phenotypes. Consequently, the model is formulated in terms of a single family environment ($C$) common to both phenotypes. As the family environment is not directly observed, an estimate of the environment, called an *index*, is created in a manner analogous to the univariate case (Laskarzewski *et al.* 1983; Rao *et al.* 1982*a*). Such an index should be carefully constructed so that it will be a measure of the family environment common to both phenotypes. Effects of assortative mating and cohabitation are incorporated in terms of correlated environments of spouses. Such a simple formulation appears to be adequate for most, if not all, physiological traits (Rao *et al.* 1979*a*; Gulbrandsen *et al.* 1979; Krieger *et al.* 1980; Morton *et al.* 1980; Rao *et al.* 1982*a*; Rao *et al.* 1982*b*). For non-physiological traits exhibiting more complex mechanisms of assortative mating, this simple formulation

might be inadequate. Intergenerational differences are retained: whereas $h_i^2$ and $c_i^2$ are the genetic and cultural heritabilities for children, they are $h_i^2 z_i^2$ and $c_i^2 y_i^2$ for adults (where the subscript $i = 1,2$ denotes the relevant phenotype). Maternal effects are also retained by distinguishing the effects of maternal ($f_M$) and paternal ($f_F$) environments on that of their children. The separate familial environments of sibs are determined partly by parental environments and partly by a non-transmitted common sibship environment ($B$).

Unlike recent univariate models (Morton, Rao & Lalouel, 1983), indices of father and mother are distinguished. The path coefficients from family environment to index are $i$ for children, $iv_F$ for fathers, and $iv_M$ for mothers. Present univariate models correspond to the constraint that $v_F = v_M = v$, an hypothesis which is testable by the likelihood ratio criterion in the present formulation.

The present model assumes that the environmental index provides a measure of transmissible environmental factors alone. That is, the model neglects possible genetic correlations between the lipoprotein variables and the index. If one or both of the lipoprotein genotypes is correlated with the index, then a reduction in the associated heritability ($h_i^2$) and an increase in the associated cultural heritability ($c_i^2$) is expected (Rao *et al.* 1982*a*). The effect of a genetic relationship between the index and the lipoproteins upon the estimation of $\rho_G$ would depend upon the nature of that relationship. If the index is associated with genetic factors common to both lipoproteins, then $\rho_G$ would be underestimated, and if associated with genetic factors unique to either lipoprotein, then $\rho_G$ would be overestimated. In either case, if the genotypes of the lipoproteins and the genotype of the index are not related, then none of the heritabilities and the genetic correlation would be affected (even if the index variables, such as obesity, are partly genetically determined).

Alternative methods of bivariate analysis have been developed by Reeve (1952), Colletto *et al.* (1981), by Hanis (1981) and Hanis & Sing (1981), by Eaves & Gale (1974), and by Lange, Boehnke & Spence (1983). These models and the present model can be distinguished in terms of (*a*) their treatment of environmental effects and cultural transmission, (*b*) incorporation of effects due to marital resemblance, and (*c*) the data set necessary to fully identify the parameters of the model.

Reeve (1952) is perhaps the first researcher to systematically study bivariate transmission using path analytic techniques. In the model he developed, each phenotype is assumed to be an additive function of genetic effects and residual environmental effects. As Reeve's primary interest was in studying transmission in fruit-flies, neither cultural transmission nor assortative mating was incorporated in the model. A phenotypic correlation is the result of either a correlation between the two genotypes or a correlation between residual environmental components or both. This later notion of residual environmental association, which was lost in many multivariate formulations which followed Reeve, can have a substantial effect upon the results of a bivariate analysis as will be seen in a later section.

Hanis (1981) and Hanis & Sing (1981) have described an eight-parameter longitudinal model for the relationship between two phenotypes. Identification of the model requires the observation of the two phenotypes at each of two times for relatives of at least two different degrees of genetic relationship. The association between the two phenotypes is modelled in terms of a correlation between the two

genotypes, an effect of common environmental factors, and a direct (regression) effect of one phenotype upon the other. Although the model does not allow for marital resemblance or cultural transmission, this, perhaps, reflects the fact that it has only been applied to the analysis of phenotypic associations for collateral relatives (sibs and cousins).

Lange *et al.* (1983) have recently extended their univariate methods of variance components analysis of pedigree data (Lange *et al.* 1976), to allow for the treatment of multiple phenotypes. As in the univariate case, the model distinguishes between dominance and additive genetic effects. The environment is modeled in terms of 'spheres of environmental influence', which, although allowing for a great degree of generality in the representation of environmental effects, do not represent a model for cultural transmission. Finally, the model makes no allowance for marital resemblance nor its effects upon phenotypic associations.

Eaves (Eaves & Gale, 1974; Martin & Eaves, 1977) as well as Plomin & DeFries (1979) have developed and applied multivariate methods to analyse twin data. The major emphasis in these analyses is to compare the mean square between and within matrices for MZ and DZ twins to identify significant sources of environmental and genetic association amongst a set of phenotypes. Genetic and environmental correlations are estimated for each pair of phenotypes and further analysed to determine their rank (i.e. the factor structure of both the environmental and genetic correlation matrices). As these methods are only applicable to twins, there is no modelling of cultural transmission or marital resemblance.

Colletto *et al.* (1981) have described a model for bivariate association which is a special case of the present model for the analysis of twin data. As in the present model, the Colletto *et al.* model allows for a genetic correlation, common environmental effects and requires an index of the common environment. Unlike the present model, there is no allowance for residual environmental correlation. As will be seen from the results of our own analyses, no allowance for residual environmental correlation can result in substantial overestimation of the genetic correlation between the two traits. As Colletto *et al.* only considered twin data, there is no explicit allowance for marital resemblance or cultural transmission; effects which are modeled in the present formulation.

Our model generates 40 correlations in nuclear families. Table 3 gives the expected correlations in terms of the parameters of Fig. 1.

The main concern in bivariate path analysis is to describe and test hypotheses about the relationship between two correlated phenotypes. In the bivariate path model formulated here, a correlation between two phenotypes could be the result of the common family environment, a correlation between the two underlying genotypes, a correlation between the residuals, or some combination of these. The next section describes statistical procedures used to describe and test hypotheses about this relationship.

## 5. STATISTICAL ANALYSIS

The statistical method of analysis effectively compares observed and expected correlations. As the latter are functions of the unknown parameters, the parameters are estimated so as to minimize discrepancies between observed and expected

Table 3. *Expected correlations in nuclear families*

| Equation number | Pair of variables* | Expected correlation† |
|---|---|---|
| 1 | $P1_F, P1_M$ | $c_1^2 y_1^2 u$ |
| 2 | $P2_F, P2_M$ | $c_2^2 y_2^2 u$ |
| 3 | $P1_F, P2_F$ | $h_1 z_1 h_2 z_2 \rho_G + c_1 y_1 c_2 y_2 + a \rho_R r_{1a} r_{2a}$ |
| 4 | $P1_F, P2_M$ | $c_1 y_1 c_2 y_2 u$ |
| 5 | $P1_{C1}, P1_{C2}$ | $h_1^2/2 + c_1^2 \psi$ |
| 6 | $P2_{C1}, P2_{C2}$ | $h_2^2/2 + c_2^2 \psi$ |
| 7 | $P1_{C1}, P2_{C1}$ | $h_1 h_2 \rho_G + c_1 c_2 + \rho_R r_{1c} r_{2c}$ |
| 8 | $P1_{C1}, P2_{C2}$ | $h_1 h_2 \rho_G/2 + c_1 c_2 \psi$ |
| 9 | $P1_F, P1_{C1}$ | $h_1^2 z_1/2 + c_1^2 y_1 (f_F + u f_M)$ |
| 10 | $P2_F, P1_{C1}$ | $h_1 h_2 z_2 \rho_G/2 + c_1 c_2 y_2 (f_F + u f_M)$ |
| 11 | $P1_F, P2_{C1}$ | $h_2 h_1 z_1 \rho_G/2 + c_2 c_1 y_1 (f_F + u f_M)$ |
| 12 | $P2_F, P2_{C1}$ | $h_2^2 z_2/2 + c_2^2 y_2 (f_F + u f_M)$ |
| 13 | $P1_M, P1_{C1}$ | $h_1^2 z_1/2 + c_1^2 y_1 (f_M + u f_F)$ |
| 14 | $P2_M, P1_{C1}$ | $h_1 h_2 z_2 \rho_G/2 + c_1 c_2 y_2 (f_M + u f_F)$ |
| 15 | $P1_M, P2_{C1}$ | $h_2 h_1 z_1 \rho_G/2 + c_2 c_1 y_1 (f_M + u f_F)$ |
| 16 | $P2_M, P2_{C1}$ | $h_2^2 z_2/2 + c_2^2 y_2 (f_M + u f_F)$ |
| 17 | $I_F, I_M$ | $i v_F i v_M u$ |
| 18 | $I_F, I_{C1}$ | $i^2 v_F (f_F + u f_M)$ |
| 19 | $I_M, I_{C1}$ | $i^2 v_M (f_M + u f_F)$ |
| 20 | $I_{C1}, I_{C2}$ | $i^2 \psi$ |
| 21 | $P1_F, I_F$ | $i v_F c_1 y_1$ |
| 22 | $P2_F, I_F$ | $i v_F c_2 y_2$ |
| 23 | $P1_M, I_M$ | $i v_M c_1 y_1$ |
| 24 | $P2_M, I_M$ | $i v_M c_2 y_2$ |
| 25 | $P1_F, I_M$ | $i v_M c_1 y_1 u$ |
| 26 | $P2_F, I_M$ | $i v_M c_2 y_2 u$ |
| 27 | $P1_M, I_F$ | $i v_F c_1 y_1 u$ |
| 28 | $P2_M, I_F$ | $i v_F c_2 y_2 u$ |
| 29 | $P1_F, I_{C1}$ | $i c_1 y_1 (f_F + u f_M)$ |
| 30 | $P2_F, I_{C1}$ | $i c_2 y_2 (f_F + u f_M)$ |
| 31 | $P1_M, I_{C1}$ | $i c_1 y_1 (f_M + u f_F)$ |
| 32 | $P2_M, I_{C1}$ | $i c_2 y_2 (f_M + u f_F)$ |
| 33 | $P1_{C1}, I_F$ | $i v_F c_1 (f_F + u f_M)$ |
| 34 | $P2_{C1}, I_F$ | $i v_F c_2 (f_F + u f_M)$ |
| 35 | $P1_{C1}, I_M$ | $i v_M c_1 (f_M + u f_F)$ |
| 36 | $P2_{C1}, I_M$ | $i v_M c_2 (f_M + u f_F)$ |
| 37 | $P1_{C1}, I_{C1}$ | $i c_1$ |
| 38 | $P2_{C1}, I_{C1}$ | $i c_2$ |
| 39 | $P1_{C1}, I_{C2}$ | $i c_1 \psi$ |
| 40 | $P2_{C1}, I_{C2}$ | $i c_2 \psi$ |

* Correlations for $(P1_M, P2_M)$ and $(P1_F, P2_F)$ are equal. Correlations for $(P1_F, P2_M)$ and $(P2_F, P1_M)$ are equal.

† $\psi = f_F^2 + f_M^2 + 2 f_F f_M u + b^2$.

correlations. However, because of better distributional properties, Fisher's $z$ transformations (Fisher, 1921) are used rather than correlations.

An observed correlation $r_i$, based upon a sample of size $n_i$, is first converted into its $z$ transformation

$$z_i = \tfrac{1}{2} \ln \frac{1 + r_i}{1 - r_i},$$

https://doi.org/10.1017/S0016672300021595 Published online by Cambridge University Press

which asymptotically follows a normal distribution with mean

$$\bar{z}_i = \tfrac{1}{2} \ln \frac{1+\rho_i}{1-\rho_i}$$

and approximate variance $1/n_i$, where $\rho_i$ is the corresponding expected correlation given in Table 3 ($i = 1, 2, \ldots, 40$). Therefore, assuming all 40 observed correlations to be independent, the overall log likelihood is (approximately) given by

$$\ln L = -\chi^2/2 + \text{constant},$$

where

$$\chi^2 = \sum_{i=1}^{40} n_i (z_i - \bar{z}_i)^2.$$

As the $\rho_i$'s are functions of the parameters (see Table 3) the $\chi^2$ is a function of the parameters and, thus the parameters can be estimated by minimizing $\chi^2$. The residual $\chi^2$ after estimating $k$ parameters follows a chi-square distribution with $40-k$ D.F. and can be used to test hypotheses on the parameters. The general model, in 18 parameters, is tested by the residual $\chi^2$ with $40-18 = 22$ D.F. If $\chi^2_{22}$ is the value of $\chi^2$ after estimating all 18 parameters and $\chi^2_{22+w}$ is another value after estimating $18-w$ of the 18 parameters, the other parameters being fixed under a null hypothesis, then $\chi^2_w = \chi^2_{22+w} - \chi^2_{22}$ provides the likelihood ratio test statistic for the null hypothesis on the $w$ fixed parameters.

All these methods have been implemeted in BPATHMIX, a computer program written in FORTRAN for Harris computers which fits path models to correlational data. The 40 familial correlations are estimated by repeated use of PATHMIX, another FORTRAN program developed on Harris computers for univariate path models (Morton *et al.* 1982; Rao *et al.* 1982*a*).

A fundamental assumption of this method of analysis is that the different sample correlation coefficients are independent. Clearly, estimates of correlations obtained from the same data are not independent of one another. Previous methods did incorporate correlations between correlations to account for this non-independence (Elston, 1975; Rao *et al.* 1979*a*), but the several analyses performed under both the independence and non-independence assumptions yielded essentially equivalent results (Rao *et al.* 1979*a*, *b*; Morton *et al.* 1980; Gulbrandsen *et al.* 1979; Krieger *et al.* 1980). Furthermore, a third method of statistical analysis, fitting path models directly to the family data, appears to give similar results to either of the methods mentioned above (i.e. assuming independent correlation estimates or incorporating correlations between the correlation estimates) (Rao *et al.* 1983). The simpler approach adopted here would thus seem justified.

## 6. RESULTS

Given two phenotypes and an index for each member of a nuclear family, with at least some families having two or more children, 40 correlations with distinct expectations can be generated. Pairs of the 9 variables $P1_F$, $P2_F$, $I_F$, $P1_M$, $P2_M$, $I_M$, $P1_{CI}$, $P2_{CI}$, and $I_{CI}$ generate 36 correlations. (See Table 1 for definition of the variables.) However, since the expected correlations are identical for ($P1_F$, $P2_F$)

and ($P1_M$, $P2_M$), and also for ($P1_F$, $P2_M$) and ($P2_F$, $P1_M$), they are pooled, reducing the number to 34. Multiple sibs generate an additional six correlations: ($P1_{C1}$, $P1_{C2}$), ($P2_{C1}$, $P2_{C2}$), ($P1_{C1}$, $P2_{C2}$), ($I_{C1}$, $I_{C2}$), ($P1_{C1}$, $I_{C2}$), and ($P2_{C1}$, $I_{C2}$) making the total number of correlations 40. These 40 correlations are estimated by the method of maximum likelihood using the computer program PATHMIX which accepts only two variables at a time on each member of a family. For each of the 3 pairs of *P1*, *P2*, and *I*, PATHMIX is executed once to obtain the relevant correlation estimates and sample sizes. For the method of calculating sample sizes, as implemented in PATHMIX, see Morton *et al.* (1983) and Rao *et al.* (1982*a*). The three bivariate analyses of lipoprotein concentrations will now be discussed separately. Observed correlations and sample sizes used in each of the 3 analyses are presented in Table 4.

### (i) *Analysis of HDL cholesterol and LDL cholesterol*

Bivariate analysis of HDL and LDL is of clinical significance because of their independent contributions to CHD risk, and because an inverse relation between the two phenotypes, especially in the upper quartile of LDL, has been observed (Khoury *et al.* 1980; Morrison *et al.* 1980). A very large negative value of $\rho_G$ is expected if this inverse relation is due to a common genetic factor.

For the purpose of data analysis, because HDL and LDL are negatively correlated, *P1* was set equal to the negation of the age–sex–contraceptive adjusted HDL value, *P2* was set equal to the age–sex–contraceptive adjusted LDL value, and *I* was set equal to $I_{\mathrm{LDL}} - I_{\mathrm{HDL}}$, where $I_{\mathrm{LDL}}$ and $I_{\mathrm{HDL}}$ are the environmental indices created separately for LDL and HDL. The latter corresponds to taking the first principal component of the two univariate indices. The effect of the sign change is to change the sign of the estimates of $\rho_G$ and $\rho_R$ which are readjusted in the analysis presented here in order to preserve the natural relationships between the variables. The 40 observed correlations ($r$) and sample sizes ($n$) are given in Table 4. Analysis of these correlations is presented in Table 5.

The analysis involves the fitting of the general model and a series of reduced models which allow tests of the hypotheses of no genetic relationship ($\rho_G = 0$), a single common genotype ($\rho_G = -1$), no intergenerational differences in residual correlation ($a = 1$), no adult residual correlation ($a = 0$), no adult or child residual correlation ($\rho_R = 0$), no genetic or residual correlation ($\rho_G = \rho_R = 0$), no intergenerational differences in heritabilities ($y_1 = y_2 = z_1 = z_2 = 1$), no assortative mating ($u = 0$) and no maternal effects ($f_F = f_M$, $v_F = v_M$). The results of these analyses for HDL and LDL are given in Table 5.

For HDL and LDL the general model provides an excellent fit to the observed correlations ($\chi^2_{22} = 17{\cdot}09$, $P = 0{\cdot}76$). The estimated genetic correlation under the general model is quite small, $\rho_G = -0{\cdot}155$, and suggests little if any genetic association between the two lipoproteins. There would also appear to be little evidence for a strong association between the residual environmental components for either children ($\rho_R$ estimated to be 0·277) or adults ($a\rho_R$ estimated to be 0). Consistent with these observations are the results of the hypothesis tests. The hypothesis of common genetic background is firmly rejected ($\chi^2_1 = 47{\cdot}40$, $P \doteq 0$), whereas we are unable to reject the hypotheses of no genetic association ($\chi^2_1 = 2{\cdot}19$,

Table 4. *Observed correlations (r) and associated sample sizes (n) for the three bivariate analyses*

| Equation number | Pair of variables | $P1$ = HDL, $P2$ = LDL* | | $P1$ = HDL, $P2$ = VLDL* | | $P1$ = LDL, $P2$ = VLDl† | |
|---|---|---|---|---|---|---|---|
| | | *r* | *n* | *r* | *n* | *r* | *n* |
| 1 | $P1_F, P1_M$ | 0·0197 | 86 | 0·0072 | 82 | 0·0141 | 94 |
| 2 | $P2_F, P2_M$ | 0·0037 | 96 | 0·0588 | 95 | 0·0589 | 95 |
| 3 | $P1_F, P2_F$ | 0·1767 | 228 | 0·4064 | 227 | 0·2423 | 225 |
| 4 | $P1_F, P2_M$ | 0·0297 | 182 | 0·0461 | 177 | −0·0166 | 188 |
| 5 | $P1_{C1}, P1_{C2}$ | 0·2998 | 163 | 0·2962 | 161 | 0·3655 | 184 |
| 6 | $P2_{C1}, P2_{C2}$ | 0·3935 | 181 | 0·3152 | 179 | 0·3546 | 155 |
| 7 | $P1_{C1}, P2_{C1}$ | 0·0902 | 273 | 0·3457 | 283 | 0·2726 | 269 |
| 8 | $P1_{C1}, P2_{C2}$ | 0·0999 | 288 | 0·1266 | 271 | 0·2104 | 272 |
| 9 | $P1_F, P1_{C1}$ | 0·1902 | 146 | 0·2263 | 142 | 0·3903 | 136 |
| 10 | $P2_F, P1_{C1}$ | 0·0066 | 153 | 0·1182 | 135 | 0·1047 | 99 |
| 11 | $P1_F, P2_{C1}$ | 0·0040 | 100 | 0·0641 | 61 | 0·0140 | 66 |
| 12 | $P2_F, P2_{C1}$ | 0·3998 | 139 | 0·2710 | 81 | 0·2812 | 76 |
| 13 | $P1_M, P1_{C1}$ | 0·2888 | 205 | 0·2890 | 204 | 0·2561 | 151 |
| 14 | $P2_M, P1_{C1}$ | 0·0901 | 190 | 0·0215 | 201 | 0·0845 | 143 |
| 15 | $P1_M, P2_{C1}$ | 0·1097 | 127 | −0·0252 | 117 | 0·0067 | 111 |
| 16 | $P2_M, P2_{C1}$ | 0·2612 | 146 | 0·0205 | 134 | 0·0173 | 123 |
| 17 | $I_F, I_M$ | 0·2085 | 83 | 0·2079 | 82 | 0·1920 | 87 |
| 18 | $I_F, I_{C1}$ | 0·2306 | 90 | 0·2211 | 91 | 0·1672 | 96 |
| 19 | $I_M, I_{C1}$ | 0·2915 | 146 | 0·2955 | 150 | 0·2689 | 150 |
| 20 | $I_{C1}, I_{C2}$ | 0·5071 | 151 | 0·5005 | 156 | 0·5147 | 160 |
| 21 | $P1_F, I_F$ | 0·2593 | 98 | 0·2306 | 94 | 0·1826 | 94 |
| 22 | $P2_F, I_F$ | 0·1515 | 98 | 0·3544 | 94 | 0·3944 | 94 |
| 23 | $P1_M, I_M$ | 0·4045 | 130 | 0·4081 | 131 | 0·2611 | 130 |
| 24 | $P2_M, I_M$ | 0·2362 | 130 | 0·4554 | 131 | 0·4602 | 130 |
| 25 | $P1_F, I_M$ | 0·0133 | 98 | 0·0388 | 94 | 0·1135 | 94 |
| 26 | $P2_F, I_M$ | 0·1123 | 98 | −0·0065 | 94 | −0·0401 | 94 |
| 27 | $P1_M, I_F$ | 0·2241 | 98 | 0·2075 | 94 | 0·0993 | 94 |
| 28 | $P2_M, I_F$ | 0·0510 | 98 | 0·0754 | 94 | 0·1006 | 94 |
| 29 | $P1_F, I_{C1}$ | −0·0152 | 85 | −0·0215 | 84 | 0·1434 | 100 |
| 30 | $P2_F, I_{C1}$ | 0·1566 | 97 | 0·1030 | 86 | 0·1013 | 91 |
| 31 | $P1_M, I_{C1}$ | 0·0375 | 130 | 0·0447 | 135 | −0·0804 | 138 |
| 32 | $P2_M, I_{C1}$ | −0·0841 | 132 | −0·0068 | 145 | −0·0196 | 146 |
| 33 | $P1_{C1}, I_F$ | 0·1203 | 140 | 0·1314 | 138 | 0·0510 | 102 |
| 34 | $P2_{C1}, I_F$ | 0·0563 | 94 | 0·1177 | 64 | 0·0928 | 64 |
| 35 | $P1_{C1}, I_M$ | 0·1913 | 203 | 0·1926 | 202 | 0·1639 | 146 |
| 36 | $P2_{C1}, I_M$ | 0·1901 | 144 | 0·0250 | 126 | 0·0237 | 116 |
| 37 | $P1_{C1}, I_{C1}$ | 0·2596 | 258 | 0·2612 | 260 | 0·2517 | 253 |
| 38 | $P2_{C1}, I_{C1}$ | 0·2735 | 243 | 0·3390 | 258 | 0·3524 | 254 |
| 39 | $P1_{C1}, I_{C2}$ | 0·1182 | 245 | 0·1220 | 248 | 0·1574 | 255 |
| 40 | $P2_{C1}, I_{C2}$ | 0·1723 | 239 | 0·1736 | 250 | 0·2109 | 242 |

* Correlations are given in terms of the negation of the adjusted HDL value. The index is created by subtracting the index for HDL from the other index.

† The index is created by adding the two separate indices for LDL and VLDL.

$P = 0{\cdot}14$), no residual association ($\chi^2_2 = 1{\cdot}27$, $P = 0{\cdot}53$) and no genetic or residual correlation ($\chi^2_3 = 2{\cdot}27$, $P = 0{\cdot}52$). Consequently the phenotypic association between HDL and LDL can be explained in terms common family environmental effects alone.

With respect to the other parameters of the model we find no evidence for intergenerational differences in heritabilities ($\chi^2_4 = 3{\cdot}92$, $P = 0{\cdot}42$) or maternal

https://doi.org/10.1017/S0016672300021595 Published online by Cambridge University Press

Table 5. *Bivariate analysis of HDL (= P1) and LDL (= P2): tests of hypotheses and estimates of some parameters*

| Hypothesis | Residual $\chi^2$ | D.F. | HDL $h_1^2$ | HDL $c_1^2$ | LDL $h_2^2$ | LDL $c_2^2$ | $u$ | $b$ | $f_F$ | $f_M$ | $\rho_G$ | $\rho_R$ | $a$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| General | 17·09 | 22 | 0·463 | 0·093 | 0·637 | 0·107 | 0·289 | 0·705 | 0·218 | 0·302 | −0·155 | 0·277 | 0 |
| $\rho_G = 0$ | 19·28 | 23 | 0·454 | 0·101 | 0·619 | 0·116 | 0·293 | 0·724 | 0·246 | 0·307 | 0 | 0·053 | 0 |
| $\rho_G = -1$ | 64·49 | 23 | 0·329 | 0·157 | 0 | 0·311 | 0·325 | 0·690 | 0·380 | 0·510 | −1 | 0·364 | 0·038 |
| $a = 1$ | 18·29 | 23 | 0·409 | 0·091 | 0·588 | 0·106 | 0·289 | 0·725 | 0·217 | 0·303 | −0·147 | 0·131 | 1 |
| $a = 0$ | 17·09 | 23 | 0·463 | 0·093 | 0·637 | 0·107 | 0·289 | 0·705 | 0·217 | 0·302 | −0·155 | 0·277 | 0 |
| $\rho_R = 0$ | 18·36 | 24 | 0·455 | 0·088 | 0·631 | 0·102 | 0·291 | 0·716 | 0·230 | 0·302 | −0·083 | 0 | 1 |
| $\rho_G = \rho_R = 0$ | 19·36 | 25 | 0·458 | 0·098 | 0·624 | 0·113 | 0·293 | 0·723 | 0·247 | 0·306 | 0 | 0 | 1 |
| $y_1 = y_2 = z_1 = z_2 = 1$ | 21·01 | 26 | 0·459 | 0·097 | 0·658 | 0·070 | 0·293 | 0·613 | 0·207 | 0·245 | −0·161 | 0·233 | 0 |
| $u = 0$ | 24·38 | 23 | 0·463 | 0·092 | 0·637 | 0·107 | 0 | 0·665 | 0·314 | 0·365 | −0·155 | 0·277 | 0 |
| $f_F = f_M$, $v_F = v_M$ | 18·64 | 24 | 0·466 | 0·090 | 0·643 | 0·104 | 0·249 | 0·701 | 0·260 | 0·260 | −0·170 | 0·296 | 0 |
| Most parsimonious: | | | | | | | | | | | | | |
| $y_1 = y_2 = z_1 = z_2 = 1$, | 29·65 | 30 | 0·445 | 0·098 | 0·648 | 0·069 | 0 | 0·678 | 0·323 | 0·323 | −0·132 | 0·065 | 1 |
| $u = 0$, $f_F = f_M$, | | | ±0·090 | ±0·029 | ±0·087 | ±0·026 | | ±0·118 | ±0·089 | ±0·089 | ±0·136 | ±0·230 | |
| $v_F = v_M$, $a = 1$ | | | | | | | | | | | | | |

Table 6. *Bivariate analysis of HDL (= P1) and* VLDL *(= P2): tests of hypotheses and estimates of some parameters*

| Hypothesis | Residual $\chi^2$ | D.F. | HDL $h_1^2$ | HDL $c_1^2$ | VLDL $h_2^2$ | VLDL $c_2^2$ | $u$ | $b$ | $f_F$ | $f_M$ | $\rho_G$ | $\rho_R$ | $a$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| General | 18·07 | 22 | 0·496 | 0·093 | 0·478 | 0·129 | 0·229 | 0·669 | 0·265 | 0·210 | −0·130 | −0·430 | 0·929 |
| $\rho_G = 0$ | 18·40 | 23 | 0·467 | 0·101 | 0·460 | 0·137 | 0·229 | 0·686 | 0·279 | 0·215 | 0 | −0·546 | 0·811 |
| $\rho_G = -1$ | 26·57 | 23 | 0·426 | 0·107 | 0·001 | 0·210 | 0·229 | 0·766 | 0·333 | 0·220 | −1 | −0·319 | 1·25 |
| $a = 1$ | 18·08 | 23 | 0·491 | 0·094 | 0·474 | 0·129 | 0·229 | 0·669 | 0·264 | 0·210 | −0·143 | −0·405 | 1 |
| $a = 0$ | 24·94 | 23 | 0·437 | 0·085 | 0·360 | 0·126 | 0·221 | 0·678 | 0·234 | 0·183 | −0·413 | −0·159 | 0 |
| $\rho_R = 0$ | 25·67 | 24 | 0·482 | 0·083 | 0·404 | 0·123 | 0·222 | 0·659 | 0·227 | 0·180 | −0·498 | 0 | 1 |
| $\rho_G = \rho_R = 0$ | 39·88 | 25 | 0·359 | 0·176 | 0·328 | 0·228 | 0·197 | 0·719 | 0·296 | 0·192 | 0 | 0 | 1 |
| $y_1 = y_2 = z_1 = z_2 = 1$ | 21·05 | 26 | 0·456 | 0·111 | 0·269 | 0·164 | 0·231 | 0·718 | 0·271 | 0·216 | −0·081 | −0·368 | 1·33 |
| $u = 0$ | 23·89 | 23 | 0·495 | 0·094 | 0·477 | 0·129 | 0 | 0·641 | 0·325 | 0·269 | −0·126 | −0·433 | 0·913 |
| $f_F = f_M$, $v_F = v_M$ | 19·12 | 24 | 0·499 | 0·092 | 0·481 | 0·127 | 0·211 | 0·675 | 0·226 | 0·226 | −0·142 | −0·420 | 1·00 |
| Most parismonious: | | | | | | | | | | | | | |
| $y_1 = y_2 = z_1 = z_2 = 1$, | 28·63 | 30 | 0·468 | 0·101 | 0·282 | 0·154 | 0 | 0·701 | 0·280 | 0·280 | −0·111 | −0·421 | 1 |
| $u = 0$, $f_F = f_M$, | | | ±0·088 | ±0·030 | ±0·109 | ±0·041 | | ±0·095 | ±0·070 | ±0·070 | ±0·214 | ±0·172 | |
| $v_F = v_M$, $a = 1$ | | | | | | | | | | | | | |

effects ($\chi^2_2 = 1{\cdot}55$, $P = 0{\cdot}46$). Although there is some evidence for assortative mating ($\chi^2_1 = 7{\cdot}29$, $P = 0{\cdot}007$), a parsimonious model which fixes $u$ at zero does give a good fit of the model to the data ($\chi^2_{30} = 29{\cdot}65$, $P = 0{\cdot}48$).

### (ii) *Analysis of HDL cholesterol and VLDL cholesterol*

These two variables are so inexplicably related that their bivariate analysis may provide valuable insights into their co-segregation. In a vast majority of cases low HDL is associated with elevated levels of VLDL.

As in the HDL–LDL analysis, *P1* was defined as the negation of the adjusted HDL level, *P2* was defined as the adjusted VLDL level, and $I$ was set equal to $I_{\text{VLDL}} - I_{\text{HDL}}$. The observed familial correlations along with their associated sample sizes are presented in Table 4; the results of the bivariate path analysis are given in Table 6. As before, the sign of the estimates of $\rho_G$ and $\rho_R$ has been adjusted to reflect the natural relationship between the two lipoproteins.

Once again there is an excellent fit of the general model to the data ($\chi^2_{22} = 18{\cdot}07$, $P = 0{\cdot}70$), with the genetic correlation estimated to be small (estimated $\rho_G = -0{\cdot}130$) and the residual environmental correlation estimated to be moderate (estimated $\rho_R = -0{\cdot}430$). The hypothesis of a common genetic background can be rejected ($\chi^2_1 = 8{\cdot}50, P = 0{\cdot}004$) while the hypothesis of no genetic association cannot be rejected ($\chi^2_1 = 0{\cdot}33$, $P = 0{\cdot}57$). There is no evidence for intergenerational differences in the magnitude of the residual environmental correlation ($\chi^2_1 = 0{\cdot}01$, $P = 0{\cdot}92$), although strong evidence that this correlation is different from zero for adults ($\chi^2_1 = 6{\cdot}87$, $P = 0{\cdot}009$) and for both adults and children ($\chi^2_2 = 7{\cdot}60$, $P = 0{\cdot}022$). Consequently, unlike the previous analysis, the phenotypic association between HDL and VLDL cannot be explained in terms of common family environmental effects only ($\chi^2_3 = 21{\cdot}81$, $P \doteq 0$).

Again we find no evidence for intergenerational differences in heritabilities ($\chi^2_4 = 2{\cdot}98$, $P = 0{\cdot}56$) and maternal effects ($\chi^2_2 = 1{\cdot}05$, $P = 0{\cdot}59$). There is evidence for some marital association ($\chi^2_1 = 5{\cdot}82$, $P = 0{\cdot}016$), but a parsimonious model which fixes marital resemblance at zero does provide for a good fit to the data ($\chi^2_{30} = 28{\cdot}63$, $P = 0{\cdot}54$).

### (iii) *Analysis of LDL cholesterol and VLDL cholesterol*

Concurrently elevated levels of both LDL and VLDL represent important risk factors for atherosclerotic disease. Consequently, an analysis of their bivariate association is of clinical significance.

In this case *P1* was defined as the adjusted LDL value, *P2* was defined as the adjusted VLDL value, and $I$ was taken as $I_{\text{LDL}} + I_{\text{VLDL}}$. The familial correlations used in the analysis are given in Table 4 while the results of the bivariate analysis are presented in Table 7.

Again there is an excellent fit of the general model to the data ($\chi^2_{22} = 18{\cdot}03$, $P = 0{\cdot}70$) with the genetic correlation estimated to be moderate (estimated $\rho_G = 0{\cdot}279$) and the residual correlation estimated to be near zero (estimated $\rho_R = 0{\cdot}010$). As in the other analyses we are able to reject the hypothesis of a single

Table 7. *Bivariate analysis of LDL (= P1) and VLDL (= P2): tests of hypotheses and estimates of some parameters*

| Hypothesis | Residual $\chi^2$ | D.F. | LDL $h_1^2$ | LDL $c_1^2$ | VLDL $h_2^2$ | VLDL $c_2^2$ | $u$ | $b$ | $f_F$ | $f_M$ | $\rho_G$ | $\rho_R$ | $a$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| General | 18·03 | 22 | 0·636 | 0·091 | 0·490 | 0·156 | 0·195 | 0·749 | 0·261 | 0·165 | 0·279 | 0·010 | 15 |
| $\rho_G = 0$ | 20·42 | 23 | 0·554 | 0·115 | 0·429 | 0·181 | 0·197 | 0·795 | 0·298 | 0·174 | 0 | 0·357 | 0·760 |
| $\rho_G = 1$ | 24·74 | 23 | 0·602 | 0·092 | 0·057 | 0·221 | 0·196 | 0·824 | 0·296 | 0·166 | 1 | −0·116 | 0 |
| $a = 1$ | 18·34 | 23 | 0·634 | 0·091 | 0·488 | 0·156 | 0·195 | 0·754 | 0·266 | 0·163 | 0·235 | 0·129 | 1 |
| $a = 0$ | 18·62 | 23 | 0·642 | 0·088 | 0·468 | 0·155 | 0·195 | 0·744 | 0·255 | 0·154 | 0·373 | −0·152 | 0 |
| $\rho_R = 0$ | 18·83 | 24 | 0·625 | 0·089 | 0·453 | 0·157 | 0·195 | 0·755 | 0·261 | 0·154 | 0·316 | 0 | 1 |
| $\rho_G = \rho_R = 0$ | 26·51 | 25 | 0·492 | 0·152 | 0·353 | 0·227 | 0·196 | 0·789 | 0·345 | 0·150 | 0 | 0 | 1 |
| $y_1 = y_2 = z_1 = z_2 = 1$ | 20·98 | 26 | 0·606 | 0·081 | 0·279 | 0·191 | 0·198 | 0·801 | 0·265 | 0·147 | 0·344 | −0·004 | 15 |
| $u = 0$ | 22·17 | 23 | 0·635 | 0·092 | 0·489 | 0·157 | 0 | 0·731 | 0·308 | 0·215 | 0·278 | 0·010 | 15 |
| $f_F = f_M$, $v_F = v_M$ | 18·65 | 24 | 0·638 | 0·090 | 0·494 | 0·154 | 0·181 | 0·759 | 0·196 | 0·196 | 0·282 | 0·011 | 15 |
| Most parsimonious: | | | | | | | | | | | | | |
| $y_1 = y_2 = z_1 = z_2 = 1$, $u = 0$, $f_F = f_M$, $v_F = v_M$, $a = 1$ | 26·63 | 30 | 0·616 ±0·086 | 0·073 ±0·026 | 0·295 ±0·117 | 0·178 ±0·046 | 0 | 0·800 ±0·099 | 0·221 ±0·067 | 0·221 ±0·067 | 0·330 ±0·192 | 0·010 ±0·217 | 1 |

common genetic background for LDL and VLDL ($\chi^2_1 = 6{\cdot}71$, $P = 0{\cdot}010$), but unable to reject the hypothesis of no genetic association ($\chi^2_1 = 2{\cdot}39$, $P = 0{\cdot}12$). There is little evidence that the residual correlation varies from generation to generation ($\chi^2_1 = 0{\cdot}31$, $P = 0{\cdot}58$), and is non-zero for either adults ($\chi^2_1 = 0{\cdot}59$, $P = 0{\cdot}44$) or for both adults and children ($\chi^2_2 = 0{\cdot}80$, $P = 0{\cdot}67$). Although the model can be fit with either no genetic correlation or no residual correlation, setting both correlations simultaneously to zero does significantly increase the residual $\chi^2$ statistic ($\chi^2_3 = 8{\cdot}48$, $P = 0{\cdot}037$). Consequently, although the results of the hypothesis tests indicate that the phenotypic association between LDL and VLDL cannot be accounted for by common family environmental effects alone, the tests are unable to resolve whether the additional source of association is due to a genetic relationship or a residual environmental relationship. Despite the equivocal nature of the hypothesis tests, when we simultaneously estimate both correlations, the maximum likelihood estimate of the genetic correlation is moderate while the maximum likelihood estimate of the residual correlation is near zero suggesting that the genetic association is the important additional source of relationship between the two lipoproteins.

Finally, we do not find evidence for intergeneration differences in heritabilities ($\chi^2_4 = 2{\cdot}95$, $P = 0{\cdot}57$) nor maternal effects ($\chi^2_2 = 0{\cdot}62$, $P = 0{\cdot}73$). We do find some evidence for marital resemblance ($\chi^2_1 = 4{\cdot}14$, $P = 0{\cdot}042$), but are able to fit a parsimonious model with no generational differences, maternal effects or marital resemblance ($\chi^2_{30} = 26{\cdot}63$, $P = 0{\cdot}64$).

## 7. DISCUSSION

In bivariate path analysis the emphasis is upon drawing inferences about the nature and source of observed phenotypic associations. In the present set of analyses, we were primarily concerned with determining whether the correlation between two lipoprotein levels could, in part, be a result of a single common genotype, two separate but correlated genotypes, correlated residual environments, or the result of common family environmental effects alone. For the three lipoproteins studied, the observed phenotypic correlations along with the maximum-likelihood estimates of the genotypic and residual environmental correlations under the parsimonious model are given in Table 8.

Although in all three analyses we were able to reject the hypothesis of a single common genotype, the results certainly were not uniform across the three pairs of lipoproteins. Furthermore, the results from the various analyses are consistent with what is known about the precursor–product relationships between the lipoproteins (Lippel *et al.* 1981; Morrison *et al.* 1982*a*, *b*). In fact, one of our primary motivations for using lipoprotein data in this first application of the bivariate model is that known metabolic relationships provide us with an independent assessment of the validity of the present bivariate formulation.

Whereas there are metabolic precursor–product relationships between VLDL and HDL (Lippel *et al.* 1981), and between VLDL and LDL (Lippel *et al.* 1981; Morrison *et al.* 1982*a*, *b*), the relationships between LDL and HDL are much less direct. Consequently, we would not expect a substantial genetic correlation

Table 8. *Phenotypic, genotypic and residual environmental correlations between the lipoproteins**

| | Phenotypic | | | Genotypic | | | Residual environmental | | |
|---|---|---|---|---|---|---|---|---|---|
| | HDL | LDL | VLDL | HDL | LDL | VLDL | HDL | LDL | VLDL |
| LDL | −0·101 | — | — | −0·132 | — | — | 0·065 | — | — |
| VLDL | −0·535 | 0·338 | — | −0·111 | 0·330 | — | −0·421 | 0·010 | — |

* Phenotypic correlations are the (sample size) weighted averages of the phenotypic correlations of adults and children, genotypic and residual environmental correlations were estimated under the parsimonious bivariate path models.

between these two lipoproteins. In fact, the estimated genetic correlation between HDL and LDL is both small ($\rho_G = -0{\cdot}132 \pm 0{\cdot}136$) and not significantly different from zero. Similarly, the residual environmental correlation is estimated to be near zero ($\rho_R = 0{\cdot}065 \pm 0{\cdot}230$) and is not significantly different from zero. The modest phenotypic association between LDL and HDL can be explained solely in terms of common family environmental effects.

The strongest phenotypic association was observed between HDL and VLDL ($r = -0{\cdot}535$). Although the estimate of the phenotypic correlation is large, the associated estimate of the genotypic correlation is small ($\rho_G = -0{\cdot}111 \pm 0{\cdot}214$) and non-significant. The important factor in the association of HDL and VLDL would appear to be a residual environmental correlation. The estimate of $\rho_R$ in the parsimonious model is $-0{\cdot}421 \pm 0{\cdot}172$, a value significantly different from zero. These results are interesting in light of the finding that elevated levels of VLDL and/or triglycerides do not independently and significantly predict risk to CHD once HDL levels have been taken into account. (Lippel *et al.* 1981). As the association between HDL and VLDL is predominantly environmental, it would appear that it is the genotype for VLDL which is not an independent predictor of CHD.

For VLDL and LDL we find both a moderate phenotypic (0·338) and genotypic ($\rho_G = 0{\cdot}330 \pm 0{\cdot}192$) correlation, with little evidence for an association between the residual environments ($\rho_R = 0{\cdot}010 \pm 0{\cdot}217$). A finding of moderate genetic association between LDL and VLDL is consistent with the known metabolic relationships between these two lipoproteins (Lippel *et al.* 1981; Morrison *et al.* 1982*a*, *b*).

The purpose of the present paper has not only been to analyse the relationship between the three lipoproteins, but also to introduce a bivariate model for the joint transmission of two phenotypes and provide some assessment of the validity of that model. It should be emphasized that the present bivariate model represents a parsimonious extension of the univariate models which, nonetheless, allows us to draw important inferences about the association between two phenotypes. The present model is much more than the simple union of two univariate models. With respect to the issue of validity, three characteristics of the present results provide support for the bivariate model used. First, in every case the $\chi^2$ statistic for the general model was less than its associated degrees of freedom (i.e. its expectation under the null hypothesis). Secondly, the results of the analysis are consistent with

Table 9. *Comparison of the maximum-likelihood estimates of the some of the parameters under the most parsimonious bivariate and univariate models**

| Lipoprotein | Parameters | | | |
|---|---|---|---|---|
| | $h^2$ | $c^2$ | $b$ | $f_F = f_M$ |
| Bivariate analyses | | | | |
| HDL | 0·457 | 0·099 | 0·692 | 0·296 |
| | ±0·089 | ±0·029 | ±0·092 | ±0·078 |
| LDL | 0·632 | 0·071 | 0·750 | 0·258 |
| | ±0·086 | ±0·026 | ±0·107 | ±0·076 |
| VLDL | 0·288 | 0·165 | 0·748 | 0·249 |
| | ±0·113 | ±0·043 | ±0·097 | ±0·068 |
| Univariate analysis | | | | |
| HDL | 0·469 | 0·124 | 0·610 | 0·360 |
| | ±0·097 | ±0·037 | ±0·138 | ±0·106 |
| LDL | 0·624 | 0·072 | 0·749 | 0·314 |
| | ±0·093 | ±0·029 | ±0·188 | ±0·131 |
| VLDL | 0·339 | 0·120 | 0·679 | 0·218 |
| | ±0·096 | ±0·036 | ±0·104 | ±0·077 |

* Univariate solution is given in Rao *et al.* (1982*a*), bivariate estimates are obtained by taking the weighted averages, the weights being given by the inverse of the variance, of the separate parameter estimates obtained in the two bivariate solutions. Standard errors for the bivariate estimates are computed using the formula for the variance of a linear combination assuming the two separate estimates are perfectly correlated.

what is known about the metabolic relationships between the lipoproteins. Finally, in every case the model which Rao *et al.* (1982*a*) found to be most parsimonious for the separate univariate analyses is the same model found to be most parsimonious in the bivariate analyses, and further, the parameter estimates obtained under the bivariate analyses are consistent with those given by the univariate analyses. Table 9 provides a comparison of the results of the bivariate and univariate analyses.

This work was supported in part by N.I.H. and N.I.M.H. grants GM 28719 and MH 31302, and by contract NO-1-HV-2-2914L from the National Heart, Lung and Blood Institute (Lipid Research Clinic's Program), General Clinical Research Center, and the CLINFO Center Grant RR-00068-19.

## REFERENCES

Colletto, G. M. D. D., Krieger, H., Magalhaes, J. R. (1981). Estimates of the genetical and environmental determinants of serum lipid and lipoprotein concentrations in Brazilian twins. *Human Heredity* **31**, 232–237.

Darlu, P., Rao, D. C., Henrotte, J. G. & Lalouel, J. M. (1982). Genetic regulation of plasma and red blood cell magnesium concentrations in man. I. Univariate and bivariate path analysis. *American Journal of Human Genetics* **34**, 874–877.

Eaves, L. J. & Gale, J. S. (1974). A method for analyzing the genetic basis of covariation. *Behavior Genetics* **4**, 253–267.

Elston, R. C. (1975). Correlations between correlations. *Biometrika* **62**, 133–148.

Gulbrandsen, C. L., Morton, N. E., Rao, D. C., Rhoads, G. G. & Kagan, A. (1979). Determinants of plasma uric acid. *Human Genetics* **50**, 307–312.

Gulbrandsen, C. L., Morton, N. E., Rhoads, G. G., Kagan, A. & Lew, R. (1977). Behavioral, social and physiological determinants of lipoprotein concentrations. *Social Biology* **24**, 289–293.

HANIS, C. L. (1981). Multivariate models for human genetic analysis: Development and application to systolic blood pressure and weight. Unpublished doctoral dissertation, University of Michigan.

HANIS, C. L. & SING, C. F. (1981). Multivariate models for human genetic analysis. I. Development of models. (In preparation.)

HEISS, G., TAMIR, I., DAVIS, C. E., TYROLER, H. A., RIFKIND, B. M., SCHONFELD, G., JACOBS, D. & FRANTZ, I. D. (1980). Lipoprotein–cholesterol distributions in selected North American populations: The Lipid Research Clinics Program Prevalence Study. *Circulation* **61**, 302–315.

KELLY, K. K., AUSTIN, M., MACIOLOWSKI, M., DAWSON, D., TYROLER, H. A., MOWERY, R. & GLUECK, C. J. (1983). The Collaborative Lipid Research Clinics Family Study: design, ascertainment, lipids and lipoproteins. *American Journal of Epidemiology* (in the Press).

KHOURY, P., MORRISON, J. A., KELLY, K. A., MELLIES, M. J., HORVITZ, R. & GLUECK, C. J. (1980). Clustering and interrelationships of coronary heart disease risk factors in school children, ages 6–19. *American Journal of Epidemiology* **112**, 524–538.

KRIEGER, H., MORTON, N. E., RAO, D. C. & AZEVEDO, E. (1980). Familial determinants of blood pressure in Northeastern Brazil. *Human Genetics* **53**, 261–266.

LANGE, K., BOEHNKE, M. & SPENCE, M. A. (1983). Extensions to pedigree analysis. V. Covariance components models for multivariate traits. *American Journal of Medical Genetics* (in the Press).

LASKARZEWSKI, P. M., RAO, D. C., MORRISON, J. A., KHOURY, P. & GLUECK, C. J. (1983). The Cincinnati Lipid Research Clinic Family Study: Social and physiological determinants of lipids and lipoprotein concentrations. *Human Heredity* (in the Press).

LIPPEL, K., TYROLER, H. A., EDER, H., GOLTO, A. & VAHOUNY, G. (1981). Relationship of hypertriglyceridemia to atherosclerosis. *Arteriosclerosis* **1**, 406–417.

MARTIN, N. G. & EAVES, L. J. (1977). The genetic analysis of covariance structure. *Heredity* **38**, 79–95.

MORRISON, J. A., KELLY, K. K., HORVITZ, R., KHOURY, P., LASKARZEWSKI, P. M., MELLIES, M. J. & GLUECK, C. J. (1982*a*). Parent–offspring and sibling lipid and lipoprotein associations during and after sharing of household environments: The Princeton School District Family Study. *Metabolism* **31**, 158–167.

MORRISON, J. A., KELLY, K. K., MELLIES, M. J., DEGROOT, I. & GLUECK, C. J. (1978). Parent–child associations at upper and lower ranges of plasma cholesterol and triglyceride. *Pediatrics* **62**, 468–478.

MORRISON, J. A., KHOURY, P., LASKARZEWSKI, P., GARTSIDE, P., MOORE, M., HEISS, G. & GLUECK, C. J. (1980). Hyperalphalipoproteinemia in hypercholesterolemic adults and children. *Transactions of Association of American Physicians* **93**, 230–243.

MORRISON, J. A., KHOURY, P., LASKARZEWSKI, P. M., MELLIES, M. J., KELLY, K., GLUECK, C. J. (1982*b*). Intrafamilial associations of lipids and lipoproteins in kindreds with hypertriglyceridemic probands: The Princeton School Family Study. *Circulation* **66**, 67–76.

MORTON, N. E., GULBRANDSEN, C. L., RAO, D. C., RHOADS, G. G. & KAGAN, A. (1980). Determinants of blood pressure in Japanese-American families. *Human Genetics* **53**, 261–266.

MORTON, N. E., RAO, D. C. & LALOUEL, J. M. (1983). *Methods in Genetic Epidemiology*. Basel, Switzerland: Karger AG.

PLOMIN, R. & DEFRIES, J. C. (1979). Multivariate behavioral genetic analysis of twin data on scholastic abilities. *Behavior Genetics* **9**, 505–517.

RAO, D. C., LASKARZEWSKI, P. M., MORRISON, J. A., KHOURY, P., KELLY, K. & GLUECK, C. J. (1982*a*). The Cincinnati Lipid Research Clinic Family Study: Cultural and Biological Determinants of Lipids and Lipoprotein Concentrations. *American Journal of Human Genetics* **34**, 888–903.

RAO, D. C., LASKARZEWSKI, P. M., MORRISON, J. A., KHOURY, P., KELLY, K. & GLUECK, C. J. (1982*b*). The Cincinnati Lipid Research Clinic Family Study: familial determinants of plasma uric acid. *Human Genetics* **60**, 257–261.

RAO, D. C., MCGUE, M., WETTE, R. & GLUECK, C. J. (1983). Path analysis in genetic epidemiology. In *Human Population Genetics: The Pittsburgh Symposium*. Stroudsburg, PA: Hutchinson Ross.

RAO, D. C., MORTON, N. E. & CLONINGER, C. R. (1979*b*). Path analysis under generalized assortative mating. I. Theory. *Genetical Research* **33**, 175–188.

RAO, D. C., MORTON, N. E., GULBRANDSEN, C. L., RHOADS, G. G., KAGAN, A. & YEE, S. (1979*a*). Cultural and biological determinants of lipoprotein concentrations. *Annals of Human Genetics* **42**, 467–477.

REEVE, E. C. R. (1952). Studies in quantitative inheritance. III. Heritability and genetic correlation in progeny tests using different mating systems. *Journal of Genetics* **51**, 520–542.

TYROLER, H. A., ANDERSON, T., CHASE, G., ELLIS, L., MOWERY, R. & VALULICK, D. (1979). The Lipid Research Clinics Population Based Family Study. In *Genetic Analysis of Common Diseases: Application to Predictive Factors in Coronary Disease* (ed. C. Sing and M. Skolnick), pp. 647–652. New York, Alan R. Liss.