

The use of measured genotype information in the analysis of quantitative phenotypes in man.

III. Simultaneous estimation of the frequencies and effects of the *apolipoprotein E* polymorphism and residual polygenetic effects on cholesterol, betalipoprotein and triglyceride levels

E. BOERWINKLE* AND C. F. SING†

* *Center for Demographic and Population Genetics, Graduate School of Biomedical Sciences, University of Texas, Houston, Texas 77225*

† *Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109-0618*

SUMMARY

We have begun a measured genotype approach to the genetic analysis of lipid and lipoprotein variability. This approach enables one to simultaneously estimate the frequencies and effects of alleles at specific loci along with the residual polygenetic variance component. In this study we consider the contribution of three common alleles at the locus coding for apolipoprotein E to interindividual variation of total cholesterol, betalipoprotein, and triglyceride levels. A sample of 102 nuclear families consisting of 434 individuals was studied. The frequencies of the $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ alleles in this sample are 0.137, 0.740, and 0.123, respectively. In separate analyses of cholesterol and betalipoprotein levels, a complete model that includes the effects of the six *apo E* genotypes, unmeasured polygenes, and individual specific environmental effects fits these data significantly better than a reduced model that does not include the effects of the *apo E* polymorphism or a reduced model that does not include the effects of polygenes. On the average the $\epsilon 2$ allele lowers total cholesterol and betalipoprotein levels by 0.425 mmol/l and 0.811 units, respectively. The $\epsilon 4$ allele is associated with an average increase of these phenotypes by 0.255 mmol/l and 0.628 units, respectively. Simultaneous estimates of the interindividual variability of total cholesterol levels attributable to the *apo E* polymorphism and to residual polygenic effects are 8% and 56%, respectively. For betalipoprotein levels, we simultaneously estimate these values to be 7% and 42%, respectively. A reduced model including the effects of polygenes but not the effects of the *apo E* polymorphism fitted the triglyceride data as well as the complete model. The estimate of the fraction of interindividual variability associated with polygenetic effects was 26.5%. We review our present understanding of the genetic architecture underlying variability of cholesterol levels in the population at large and infer that the majority of the genetic variability may be accounted for by polymorphic gene loci with moderate effects on cholesterol levels.

INTRODUCTION

Biometrical genetic studies have established that a significant proportion of the interindividual variability of many quantitative phenotypes is attributable to polygenetic differences among individuals. These studies assume that there are a large number of unlinked loci (polygenes) with small, independent and additive effects contributing to the phenotype of interest. This assumption has often been accepted as a biological fact. Findings from limited

experimental animal work suggest that this assumption may not be valid, at least for some traits (Thompson, 1975; Thompson & Thoday, 1979; Hopkinson *et al.* 1964). Their work suggests that relatively few gene loci may be responsible for the genetic component of quantitative phenotypic variability. Direct estimation of the frequencies and effects of alleles at specific loci is required to elucidate the genetic architecture (number of loci, frequency of their alleles, size of their effects and type of loci) of a quantitative phenotype.

In order to investigate the frequencies, effects, and the contribution to phenotypic and polygenetic variance of a specific locus, we have implemented a 'measured genotype' approach to the genetic analysis of a quantitative phenotype (Boerwinkle *et al.* 1986a). This approach uses measured genetic variability at loci that are physiologically involved in the determination of a quantitative phenotype. Lipid metabolism is a system that allows one to pursue a measured genotype approach to genetic variability. Biometrical genetic studies have demonstrated that a significant proportion of the variability of many lipid and lipoprotein phenotypes in the population at large is attributable to genetic differences among individuals (Sing & Orr, 1978; Rao *et al.* 1979). Applications of this approach by ourselves (Sing & Davignon, 1985; Boerwinkle *et al.* 1986b) and others (Uterman *et al.* 1979; Robertson & Cumming, 1985), have utilized samples of unrelated individuals and primarily investigated the frequencies and effects of alleles at only the measured locus.

By sampling groups of related individuals and measuring the value of a quantitative phenotype and the genotype at a single locus, one can directly partition the overall genetic variability into a proportion attributable to the measured single locus and a proportion due to the segregation of unmeasured residual polygenes. The logic and methods that enable one to simultaneously estimate the frequencies and effects of the alleles at a single measured locus, and a residual polygenetic variance component, have been presented in the first paper in this series (Boerwinkle *et al.* 1986a). In this paper, we simultaneously estimate the frequencies and effects of a single measured locus, and the residual polygenetic variance in order to begin to investigate the role of the apolipoprotein (*apo*) E polymorphism in determining the genetic architecture underlying quantitative phenotypic variability in three measures of lipid metabolism.

Many of the phenotypes that are measures of lipid metabolism, such as total serum cholesterol levels, are associated with one's relative risk of coronary artery disease (Davignon *et al.* 1983). Genetic loci involved in lipid metabolism have been identified and allelic variability at many of these loci has been described (reviewed in Breslow, 1985). In this study we are continuing our efforts to understand the genetic variability underlying lipid and lipoprotein quantitative phenotypes. We have measured total cholesterol, betalipoprotein and triglyceride levels, and scored the *apo* E polymorphism on a sample of 102 nuclear families consisting of 434 individuals. Apo E has an important role in lipid metabolism as a structural component of VLDL, a ligand for lipoprotein receptors and is involved in the conversion of VLDL to LDL by lipoprotein lipase. Apo E is also polymorphic (Utermann *et al.* 1977; Zannis & Breslow, 1981; Bouthillier *et al.* 1983). There are three common isoforms of apo E designated E2, E3 and E4 which are distinguishable by isoelectric focusing. The three alleles that code for these isoforms combine to produce six possible genotypes ($\epsilon 2/2$, $\epsilon 2/3$, $\epsilon 2/4$, $\epsilon 3/3$, $\epsilon 3/4$ and $\epsilon 4/4$). We wish to investigate the contribution of the *apo* E polymorphism to the determination of the polygenetic variability of cholesterol, betalipoprotein and triglyceride levels. These analyses are the first efforts to

directly partition the contribution of a single polymorphic locus out of the polygenetic variance for these quantitative phenotypes.

METHODS

Samples considered

The analyses reported here were carried out on a sample of 102 nuclear families obtained as part of a large population based preventive medicine study in Nancy, France (Houot *et al.* 1985). The families were sequentially sampled at the Centre for Preventive Medicine in Nancy between August 20, 1983 and September 12, 1983. These 102 families consist of 434 individuals (204 parents and 230 children). Only criteria that do not allow routine phlebotomy were reasons for exclusion. There are no individuals in this sample recorded as using antihyperlipidemic drugs. To compare adults and children for the effects of the *apo E* polymorphism on lipid and lipoprotein variability, the 204 parents and the youngest child from each family were considered.

Laboratory procedures

Blood was collected into heparin containing tubes between 8.00 a.m. and 9.00 a.m. or between 11.45 a.m. and 12.45 p.m. from fasting and supine individuals. Total plasma cholesterol was measured enzymatically on an SMA (Technicon) by the method of Allain *et al.* (1974); triglycerides were measured enzymatically on a GSA II (Greiner) by the method of Steinmetz & Panek (1978). Betalipoprotein levels were measured using a turbidimetric method (Burstein & Samaille, 1959). This method estimates lipid and lipoprotein components of LDL and VLDL, which are reported in relative units of absorbance.

A 2 ml aliquot of plasma from each subject was sent to Ann Arbor on dry ice for typing the *apo E* polymorphism by two-dimensional gel electrophoresis. Plasma samples were kept at -80°C until they were electrophoresed. Two microlitres of plasma were used for each two-dimensional gel. The two-dimensional electrophoresis was carried out as previously described (Neel *et al.* 1984; Boerwinkle *et al.* 1986*b*). First-dimension electrophoresis utilized ampholytes (LKB) with a pH range of 3.5 to 10 and was carried out for 15 h at 1200 V and 2 h at 1500 V. The slab gels contained a 110–130 g/l acrylamide gradient. Following electrophoresis, the gels were initially stained with Coomassie brilliant blue and then were subsequently silver stained by the method of Sammons *et al.* (1981). The three isoforms of *apo E* appear as separate spots on the gel. The two-dimensional electrophoretograms reported here were typed concordantly and independently by two experienced observers. Typing of a gel was done without knowledge of the individual's lipid and lipoprotein levels, or the *apo E* type of his or her relatives.

Statistical methods

Before investigating the effects of the *apo E* polymorphism on the lipid and lipoprotein phenotypes, multiple linear regression was used to adjust these traits for the effects of variation in gender, generation, hormone use in women, age within generation, height and weight. The model and analytical methods for estimating from pedigree data the frequencies and effects of the *apo E* polymorphism, the polygenetic variance component, and a nongenetic variance component were given in the first paper of this series (Boerwinkle *et al.* 1986*a*). Briefly, we model the adjusted lipid or lipoprotein level of the *i*th individuals as

$$y_{ij} = \mu_j + G_i + E_i, \quad (1)$$

where μ_j is the mean of the j th genotype at a single locus, G_i is the deviation from μ_j due to the polygenotype of the i th individual, and E_i is the deviation from μ_j due to totality of all environmental effects and measurement errors specific to the i th individual. The polygenotype is assumed to be determined by a large number of unlinked loci each with small additive allelic effects that combine independently and additively. G and E are random effects that are assumed to be independent and normally distributed, each with a mean of zero and variance σ_G^2 and σ_E^2 , respectively. The likelihood of observing the measured single locus genotypes and phenotypes of a family can be partitioned into the likelihood of observing the measured genotypes times the likelihood of the phenotypes of the family members conditional on their measured single locus genotypes. The likelihood of the measured single locus genotypes is a simple function of the genotype or allele frequencies of the measured locus and Mendelian transmission probabilities. The likelihood of the phenotypes of the members of a family, conditional on their measured single locus genotypes, is a multivariate normal distribution with mean vector μ and covariance matrix Σ . The mean vector takes the value of the genotypic means determined by the measured genotype of the family members. The elements of Σ are a function of the variance components of the model and the coefficient of relatedness between individuals (Boerwinkle *et al.* 1986a). The parameters of the model to be estimated include the genotype frequencies for the measured single locus (f_j), the genotype specific means (μ_j), the polygenetic variance component (σ_G^2), and a variance component due to individual specific non-genetic effects (σ_E^2). Estimates of the allele frequencies were obtained by allele counting using the measured genotypes of the parents. Estimates of the means and variance components were obtained using Fisher's scoring algorithm and the derivatives given previously (Boerwinkle *et al.* 1986a). To evaluate the precision of the parameter estimates in one case, we also present the estimated variance-covariance matrix of the parameter estimates for our analysis of total plasma cholesterol levels. The elements of this matrix are the elements of the negative of the inverse of the information matrix evaluated at the maximum likelihood estimates of the parameters of the model.

Hypothesis testing was carried out by comparing complete and reduced models. Parameterization of the complete model includes the frequencies and means of the six *apo E* genotypes, σ_G^2 and σ_E^2 . Reduced models were generated from the complete model by fixing one or more parameters to hypothesized values. The test statistic used to compare competing hypotheses is equal to $-2(L_R - L_C)$ where L_C is the maximum \log_e likelihood of the complete model and L_R is the maximum \log_e likelihood of the reduced model. The likelihood ratio statistic is distributed approximately as a chi-square with degrees of freedom determined by the number of parameters constrained in the reduced model. Reduced models of specific interest here are: (1) the *apo E* polymorphism has no effect on lipid and lipoprotein levels; (2) the residual polygenetic variance component is equal to zero; and (3) a subset of the *apo E* genotype specific means are equal to one another. The subsets of means set equal to one another were chosen based on the results of previous analyses on the effects of the *apo E* polymorphism on these and other phenotypes (Menzel *et al.* 1983; Robertson & Cumming, 1985; Sing & Davignon, 1985; Boerwinkle *et al.* 1986b). From these studies, we hypothesize that the mean cholesterol and betalipoprotein levels of individuals with the $\epsilon 2/2$ and $\epsilon 2/3$ genotypes were equal to one another, as are the mean levels of individuals with the $\epsilon 2/4$, $\epsilon 3/4$ and $\epsilon 4/4$ genotypes. No *a priori* hypotheses were formulated about the effects of the *apo E* polymorphism on triglyceride levels.

Table 1. *Phenotypic means and (standard deviations) for each type of family member and the relative frequencies of their apo E genotypes*

	Fathers	Mothers	Male children	Female children
No. of individuals	102	102	113	117
Age (months)	498 (86)	473 (84)	145 (56)	157 (54)
Height (cm)	171 (6)	159 (6)	146 (23)	146 (19)
Weight (kg)	72 (9)	60 (9)	40 (18)	40 (14)
Cholesterol (mmol/l)	6.27 (1.10)	5.67 (0.99)	4.74 (0.72)	4.95 (0.90)
Betalipoprotein (absorbance)	9.51 (2.85)	7.15 (2.18)	5.90 (1.29)	6.12 (1.56)
Triglycerides (mmol/l)	1.27 (0.74)	0.84 (0.48)	0.70 (0.28)	0.76 (0.33)
Apo ε2/2	0.00	0.03	0.00	0.00
Apo ε2/3	0.21	0.20	0.20	0.20
Apo ε2/4	0.02	0.05	0.02	0.02
Apo ε3/3	0.58	0.51	0.57	0.54
Apo ε3/4	0.17	0.20	0.20	0.23
Apo ε4/4	0.02	0.01	0.01	0.01

Four methods were used to estimate the contribution of the *apo E* polymorphism to the phenotypic and polygenetic variance. The expected contribution of the *apo E* polymorphism to the phenotypic variance in the population at large is

$$V(\mu_j) = \sum f_j(\mu_j - \mu.)^2, \tag{2}$$

where *j* indicates the *j*th *apo E* genotype and $\mu. = \sum f_j \mu_j$. Hence the variance of the phenotype in the population is equal to

$$V(y_{ij}) = \sum f_j(\mu_j - \mu.)^2 + \sigma_G^2 + \sigma_E^2. \tag{3}$$

The first method of estimating the variance contribution that we considered simply substitutes parameters of the model in equation (2) with parameter estimates. This estimator is biased (Boerwinkle & Sing, 1986). It will, on the average, overestimate the variance contribution of the single measured locus. For a sample of unrelated individuals, the bias correction is equal to $((J-1)/n)(\sigma_G^2 + \sigma_E^2)$, where *J* is the number of genotype classes and *n* is the number of unrelated individuals in the sample. For a sample of related individuals, the expectation of the magnitude of the bias has not been derived. We took two approaches in formulating a bias corrected estimator. We set *n* equal to the number of parents in the sample (the second method) and also set *n* equal to the total sample size (the third method). The fourth method considered estimates the variance contribution of the single locus by the increase in the estimated polygenetic variance component when the measured single locus effects are not parameterized in the model over the estimated variance component when the measured single locus effects are included in the model.

Table 2. *Maximum likelihood estimates of parameters of selected models*
 (The relative log_e likelihood for each model and tests of hypotheses are also given.)

	Models			
	Complete	No apo E effect	$\sigma_G^2 = 0$	$\mu_{22} = \mu_{23}$ μ_{33} $\mu_{24} = \mu_{34} = \mu_{44}$
Cholesterol				
μ_{22}	4.71	5.38	4.84	4.94
μ_{23}	4.94	5.38	4.94	4.94
μ_{33}	5.46	5.38	5.43	5.45
μ_{24}	5.28	5.38	5.33	5.62
μ_{34}	5.65	5.38	5.62	5.62
μ_{44}	5.91	5.38	5.97	5.62
σ_G^2	0.399	0.448	(0.00)	0.403
σ_E^2	0.309	0.324	0.732	0.311
$h^{2[1]}$	0.563	0.580	(0.00)	0.564
log _e likelihood	-520.61	-537.77	-549.29	-522.25
$\chi^{2[2]}$ (d.f.)	—	34.32** (5)	57.36** (1)	3.28(NS) (3)
Betalipoproteins				
μ_{22}	5.85	7.11	5.92	6.27
μ_{23}	6.26	7.11	6.31	6.27
μ_{33}	7.23	7.11	7.17	7.22
μ_{24}	7.01	7.11	7.16	7.64
μ_{34}	7.68	7.11	7.63	7.64
μ_{44}	8.87	7.11	8.92	7.64
σ_G^2	1.41	1.54	(0.00)	1.42
σ_E^2	1.91	2.03	3.34	1.93
$h^{2[1]}$	0.42	0.43	(0.00)	0.42
log _e likelihood	-865.70	-880.97	-879.85	-867.87
$\chi^{2[2]}$ (d.f.)	—	30.54** (5)	28.30** (1)	4.34(NS) (3)
Triglycerides				
μ_{22}	1.324	0.885	1.318	
μ_{23}	0.892	0.885	0.898	
μ_{33}	0.862	0.885	0.860	
μ_{24}	1.051	0.885	1.034	
μ_{34}	0.894	0.885	0.889	
μ_{44}	1.046	0.885	1.044	
σ_G^2	0.053	0.0546	(0.00)	
σ_E^2	0.150	0.151	0.204	
$h^{2[1]}$	0.262	0.265	(0.00)	
log _e likelihood	-265.47	-268.30	-271.96	
$\chi^{2[2]}$ (d.f.)	—	5.66(NS) (5)	12.98** (1)	

[1] $h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}$.

[2] $\chi^2 = -2(\log_e(\text{Complete}) - \log_e(\text{Reduced}))$ () indicates fixed in model to hypothesized value.

* $P < 0.05$. $P < 0.01$.

RESULTS

Descriptive statistics for the family members are given in Table 1. One 'family' had no children and the other families had an average of 2.28 children per family. As expected, the fathers of these nuclear families were older, taller, heavier, and had higher lipid and lipoprotein levels than their spouses. The mean levels of the measured concomitants, and lipid and lipoprotein phenotypes were higher for parents than for children. For children, the mean levels of the measured concomitants and the lipid and lipoprotein phenotypes were comparable between genders. 35% of the interindividual variability in plasma cholesterol levels in this sample could be accounted for by a linear relationship with generation, gender, hormone use among females, age, height and weight. This value is 41% and 24% for betalipoprotein and triglycerides, respectively. Adjustment by linear regression for variation in these concomitant variables results in a statistically significant ($P < 0.01$) reduction in variance for each of the three phenotypes.

The relative frequencies of the six *apo E* genotypes for each gender by generation group are also given in Table 1. These relative frequencies are not significantly heterogeneous among these groups. Using the genotype counts from the parents, the estimated relative frequencies of the $\epsilon 2$, $\epsilon 3$ and $\epsilon 4$ alleles in the population at large are 0.137 (± 0.00029), 0.740 (± 0.00098) and 0.123 (± 0.00080), respectively. The relative frequencies of the $\epsilon 2$, $\epsilon 3$ and $\epsilon 4$ alleles in the children are 0.111 (± 0.00022), 0.761 (± 0.00089) and 0.128 (± 0.00077), respectively. The genotype counts of the parents and the children do not differ significantly from those expected assuming Hardy-Weinberg equilibrium. Using a chi-square goodness-of-fit test (Vogel & Motulsky, 1979), the segregation of the *apo E* genotypes among the children in the families does not differ significantly from the ratios expected under a model of autosomal codominant inheritance with Mendelian segregation probabilities.

The maximum likelihood estimates of the *apo E* genotype specific means, the polygenetic variance component, and the individual specific environmental variance component for various models are given in Table 2. Table 2 also presents the \log_e likelihood of the parameter estimates of the model given the quantitative phenotypes and the *apo E* genotypes and the difference between the \log_e likelihood of the complete model and each of the reduced models.

For total plasma cholesterol levels, the complete model fitted the data significantly better than a model with no polygenes ($P < 0.01$) and a model with no effect of the *apo E* polymorphism ($P < 0.01$). The reduced model that groups genotype specific means based on *a priori* hypotheses, fitted the data as well as the complete model. Individuals with the $\epsilon 2/2$ genotype had the lowest average cholesterol in the sample and individuals with the $\epsilon 4/4$ genotype had the highest average cholesterol in the sample. Individuals with *apo E* genotypes $\epsilon 2/2$ and $\epsilon 2/3$ had lower average cholesterol levels than the grand mean, and individuals with *apo E* genotypes $\epsilon 2/4$, $\epsilon 3/4$ and $\epsilon 4/4$ had higher average cholesterol levels than the grand mean. $\epsilon 3/3$ individuals had an average cholesterol level close to the grand mean. For the complete model, 56% of the phenotypic variance within *apo E* genotype class was attributable to residual polygenetic effects. When there was no *apo E* effect parameterized in the model, the polygenetic variance contribution to the variance of cholesterol levels was 58%.

The complete model describing the distribution of betalipoprotein levels also fitted these data significantly better than a model with no polygenes ($P < 0.01$), significantly better than a model

Table 3. *Variance, covariance and correlation matrix of the parameter estimates of the full model for our analysis of cholesterol levels*

(The variances of the parameter estimates are on the diagonal of the matrix, the covariances between the parameter estimates are in the lower triangle of the matrix and the correlations between the parameter estimates are in the upper triangle of the matrix.)

θ	$\hat{\theta}$	μ_{22}	μ_{23}	μ_{33}	μ_{24}	μ_{34}	μ_{44}	σ_G^2	σ_E^2
		4.72	4.94	5.46	5.28	5.65	5.91	0.399	0.309
μ_{22}		0.0821	0.1532	0.0210	0.0651	0.0134	0.0132	0.0000	0.0000
μ_{23}		0.0043	0.0096	0.1692	0.1095	0.0786	0.0128	0.0000	0.0000
μ_{33}		0.0004	0.0011	0.0044	0.0281	0.1596	0.0190	0.0000	0.0000
μ_{24}		0.0040	0.0023	0.0004	0.0459	0.1168	0.0705	0.0000	0.0000
μ_{34}		0.0004	0.0008	0.0011	0.0026	0.0108	0.1453	0.0000	0.0000
μ_{44}		0.0012	0.0004	0.0004	0.0048	0.0048	0.1011	0.0000	0.0000
σ_G^2		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0060	-0.7307
σ_E^2		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0031	0.0030

with no effect of the *apo E* polymorphism ($P < 0.01$), but not significantly better than the reduced model based on *a priori* results of the effects of the *apo E* polymorphism on betalipoprotein levels. Individuals with the $\epsilon 2/2$ and $\epsilon 4/4$ genotypes have the lowest and highest average betalipoprotein levels, respectively. As for cholesterol, individuals carrying an $\epsilon 2$ allele tended to have lower betalipoprotein levels than the grand mean, and individuals carrying an $\epsilon 4$ allele tended to have higher betalipoprotein levels than the grand mean. 42% of the betalipoprotein variance within *apo E* genotype class was attributable to residual polygenetic effects.

For triglycerides, the complete model predicts that individuals carrying an $\epsilon 3$ allele have lower triglyceride levels than individuals not carrying an $\epsilon 3$ allele. 26% of the within *apo E* genotype phenotypic variance can be accounted for by the effects of residual polygenes. The complete model describing the distribution of triglyceride levels among family members did not fit the data significantly better than a model with no *apo E* effects, but significantly better than a model with no polygenes. Assuming no effects of the *apo E* polymorphism, a model with polygenes fitted these data significantly better than a model without polygenes (data not shown; χ^2 (1 d.f.) = 12.86, $P < 0.01$). These data predict that the *apo E* polymorphism has no statistically significant effect on triglyceride levels and that 26.5% of the variability of plasma triglyceride levels is attributable to polygenetic differences among individuals.

The estimated variances, covariances and correlations of, and between, the parameter estimates of the complete model for our analysis of cholesterol levels are given in Table 3. The average magnitude of the standard deviation of the estimates of the single locus genotype means is only 0.181. Measured genotype analyses typically yield smaller variances of the parameter estimates than unmeasured genotype analyses, such as complex segregation analysis (Boerwinkle, 1985). The variances of the variance component estimates are also small. The covariances and correlations between the estimated genotype specific means and variance components are zero because of the block diagonal structure of the information matrix (Boerwinkle *et al.* 1986a). The correlations between the estimates of the genotype specific means are all positive and low. The correlation between the estimates of the residual polygenetic and individual specific nongenetic variance components is negative and quite large. We attribute this to a fixed within *apo E* genotype phenotypic variance in the sample that can be partitioned into the two variance

Table 4. Estimates of the contribution of the apo E polymorphism to the genetic and phenotypic variance

Method	Var (apo E)	Var (apo E)	Var (apo E)
		Var (apo E) + σ_G^2	Var (apo E) + $\sigma_G^2 + \sigma_E^2$
Cholesterol			
$\Sigma f_j(\mu_j - \mu.)^2$	0.0690	0.147	0.088
$\Sigma f_j(\mu_j - \mu.)^2$ - $((J-1)/n^*)(\sigma_G^2 + \sigma_E^2)$	0.0516	0.114	0.068
$\Sigma f_j(\mu_j - \mu.)^2$ - $((J-1)/n\ddagger)(\sigma_G^2 + \sigma_E^2)$	0.0608	0.132	0.079
$\sigma_G^2 - \sigma_G^2\ddagger$	0.0490	0.109	0.065
Betalipoprotein			
$\Sigma f_j(\mu_j - \mu.)^2$	0.290	0.171	0.080
$\Sigma f_j(\mu_j - \mu.)^2$ - $((J-1)/n^*)(\sigma_G^2 + \sigma_E^2)$	0.209	0.129	0.059
$\Sigma f_j(\mu_j - \mu.)^2$ - $((J-1)/n\ddagger)(\sigma_G^2 + \sigma_E^2)$	0.252	0.152	0.070
$\sigma_G^2 - \sigma_G^2\ddagger$	0.13	0.084	0.038
Triglyceride			
$\Sigma f_j(\mu_j - \mu.)^2$	0.0052	0.089	0.0250
$\Sigma f_j(\mu_j - \mu.)^2$ - $((J-1)/n^*)(\sigma_G^2 + \sigma_E^2)$	0.0002	0.0038	0.0010
$\Sigma f_j(\mu_j - \mu.)^2$ - $((J-1)/n\ddagger)(\sigma_G^2 + \sigma_E^2)$	0.0029	0.0519	0.0141
$\sigma_G^2 - \sigma_G^2\ddagger$	0.0030	0.054	0.0146

* $n = 204$

† $n = 434$

‡ σ_G^2 is the polygenetic variance component in a model that does not parameterize the measured genotype information.

components. The sum of the two variance components is fixed for this sample – so, if one goes up, the other must go down.

The contribution of the apo E polymorphism to the variance of cholesterol, betalipoprotein and triglyceride levels is given in Table 4. This contribution has been estimated using four different estimators: substituting parameters in equation (2) with parameter estimates, two bias corrected estimators, and the difference between the estimated polygenetic variance components of the complete model and a model with no apo E effects parameterized. The proportion of the estimated genetic variance and phenotypic variance attributable to the apo E polymorphism estimated by each method are also given. The contribution of the apo E polymorphism to variability of cholesterol levels ranges from 6.5% to 8.8%, depending on the method used. We estimate that this polymorphism accounts for between 10.9% and 14.7% of the genetic variance underlying cholesterol variability, depending on the method used. The apo E polymorphism accounts for between 3.8% and 8%, and between 0.1% and 2.5% of the variability of betalipoprotein and triglyceride levels, respectively. No other loci have been identified with such a large contribution to the interindividual variability of a cardiovascular disease risk factor in the population at large.

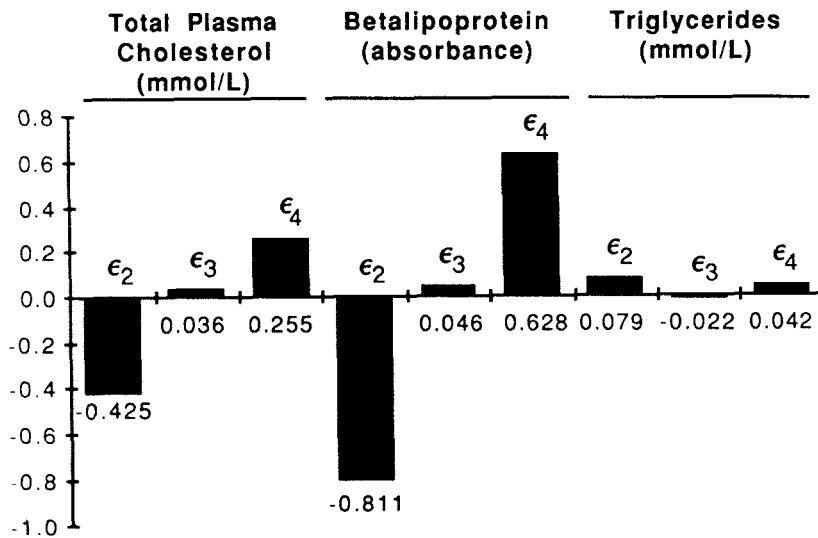


Fig. 1. Average effects of the *apo E* alleles.

DISCUSSION

Using measured genotype analyses of data from 434 individuals from 102 nuclear families, we have simultaneously estimated the frequency and effects of the *apo E* polymorphism, and the residual polygenetic effects on quantitative levels of total plasma cholesterol, betalipoproteins and triglycerides. Individuals with the $\epsilon_2/2$ and $\epsilon_2/3$ genotypes had lower average cholesterol and betalipoprotein levels than the grand mean and individuals with the $\epsilon_3/4$ and $\epsilon_4/4$ genotypes had higher average cholesterol and betalipoprotein levels than the grand mean. In general, the frequency and effects of the *apo E* polymorphism reported here are in accordance with the results of others (reviewed in Sing & Davignon, 1985; Boerwinkle *et al.* 1986b). The residual polygenetic variance accounts for 56% and 42% of the within *apo E* genotype variance of cholesterol and betalipoproteins, respectively. No effect of the *apo E* polymorphism on triglyceride levels was detected. Unmeasured polygenetic differences among individuals account for 26.5% of the interindividual variability of triglyceride levels.

With the results presented here, we have begun to investigate the genetic architecture underlying cholesterol, betalipoprotein and triglyceride variability. Using the estimated allele frequencies and the estimated genotype specific phenotypic means, we next calculated the average effects of the *apo E* alleles on total cholesterol, betalipoprotein and triglyceride levels (Fig. 1). Formulae for estimating the average effects of the *apo E* alleles have been previously presented (Sing & Davignon, 1985; Boerwinkle *et al.* 1986b). The average effect of an *apo E* allele may be interpreted as the average deviation from the sample mean of individuals carrying that allele. The average effect of the ϵ_2 allele is to lower cholesterol and betalipoprotein levels by 0.425 mmol/l and 0.811 units, respectively. The average effect of the ϵ_4 allele is to raise the level of these phenotypes by 0.255 mmol/l and 0.628 units, respectively. The average effect of the ϵ_3 allele is near zero. The average effects of the *apo E* alleles on triglyceride levels are also near zero. These average effects of the *apo E* alleles are similar to those reported for other populations (reviewed in Boerwinkle *et al.* 1986b).

The model given in equation (1) assumes that the average effects of an *apo E* allele are the

same for all individuals carrying that allele. Our previous analyses have indicated that there is no evidence for statistically significant interaction between the effects of the *apo E* polymorphism and the effects of gender (Boerwinkle *et al.* 1986*b*). The results reported here are the first time that the frequency and effects of the *apo E* polymorphism have been examined in a sample including children. We examine here whether the effects of the *apo E* polymorphism on total cholesterol levels are consistent between generations. To do this, we created two subsamples to compare. The first subsample consists of the parents of the families ($N = 204$) and the second subsample consists of the youngest child from each family ($N = 101$; recall that one 'family' was childless). Among the sample of parents, the average effects of the $\epsilon 2$, $\epsilon 3$ and $\epsilon 4$ alleles are -0.54 mmol/l, 0.06 mmol/l and 0.24 mmol/l, respectively. Among the sample of children, the average effects of these alleles are -0.67 , 0.06 and 0.31 mmol/l, respectively. The distributional properties of the estimated vector of average effects of alleles are unknown and a method for testing the equality of a vector of average effects of alleles is not obvious. We used a sample-resample method, described in the appendix, to test the hypothesis that the average effects of the *apo E* alleles are equal between generations. We infer that the average effects of the *apo E* alleles are not statistically different between the parents and the youngest child of the families.

The variability among estimates of the contribution of the *apo E* polymorphism to the phenotypic variance presented in Table 4 is not large. For total plasma cholesterol levels, the contribution of the *apo E* polymorphism to the phenotypic variance ranged from 6.5% to 8.8%. It is not clear which estimator is more accurate. We know that the estimator derived directly from the parametric definition given in equation (2) is biased upward (Boerwinkle & Sing, 1986). The two bias corrections applied here are intuitive and may not be the correct magnitude. The form of the bias correction is based on the bias correction appropriate for a sample of unrelated individuals (Boerwinkle & Sing, 1986). The estimator based on the difference between polygenetic variance components may also be biased. Clearly, the problem of estimating the contribution of a single polymorphic locus to the determination of the polygenetic variance has yet to be adequately resolved.

One advantage of the measured genotype approach to quantitative phenotypic variability is that it creates a more complete picture of the sources of genetic variability contributing to the phenotype. A measured genotype approach yields critical information for elucidating the genetic architecture of quantitative phenotypic variability. Below, we summarize our understanding about the genetic architecture of total cholesterol variability. This understanding is developed from the work presented here, in the other papers in this series, and by the work of others. Ignoring the measured genotype information, we report here that 58% of the interindividual variability in adjusted cholesterol levels is attributable to genetic differences among individuals. This number is similar to that reported by others (Sing & Orr, 1978; Rao *et al.* 1979). It is the goal of a measured genotype approach to partition this polygenetic variance contribution into the contribution of individual measured loci and residual unmeasured polygenes. In the present study, we estimate that about 7% of the variability of cholesterol levels among individuals is attributable to the *apo E* polymorphism. We also estimate that 56% of the within *apo E* genotype cholesterol variability is attributable to residual polygenetic effects. In earlier studies (Boerwinkle *et al.* 1986*b*; Boerwinkle *et al.* 1987) using samples of unrelated adults, we reported that the *apo E* and *apo A-IV* polymorphisms account for 8.6%

and 6.6%, respectively, of the interindividual variability in adjusted total cholesterol levels. Talmud & Humphries (1986) described an Xba I restriction fragment length polymorphism in the *apo B* gene that accounts for 18% of the cholesterol variability in a sample of 61 individuals from London. Moll *et al.* (1986) have reported evidence from segregation analysis of the existence of a polymorphic locus with a major effect on apolipoprotein A-I levels that accounts for 1% of the variance of cholesterol levels. Sing & Orr (1976) identified 4 polymorphic blood markers (*Gm. haptoglobin*, *Se* and *ABO*) that had a small but significant contribution (about 1%) to the interindividual variability of cholesterol levels among 6366 individuals from Tecumseh, Michigan. Genetic variability at the *LDL receptor* locus (Goldstein & Brown, 1979) accounts for less than 1% of the cholesterol variability in the general population (Sing *et al.* 1975). This is not a complete list. There are likely other polymorphic loci whose products are involved in lipid metabolism, the effects of which are not considered here. Assuming that there is no statistical interaction between the effects of these loci, the combined contribution of these loci to the interindividual variability of cholesterol levels is approximately 34%.

At least two major points emerge from these studies. The first major point is that if the results reviewed here give any indication of future directions, as more measured genotype analyses are carried out on a given phenotype, the sum of the contributions of individual loci to the phenotypic variance may exceed the polygenetic variance component. The few loci discussed above account for approximately 34% of the phenotypic variance for total serum cholesterol levels. Assuming that 58% of the phenotypic variance can be accounted for by polygenetic differences among individuals, these loci account for 59% of the polygenetic variance. If the other loci have a contribution similar to the *apo E* and *apo A-IV* polymorphisms, it will only require four or five more to account for all the genetic variability underlying cholesterol variability. How can this be so in light of the fact that there are likely many loci affecting cholesterol levels? We believe there are four likely explanations. The first is that the samples reviewed above are heterogeneous in such a way that the results cannot be pooled across studies. Although this is likely the case, we believe that these few loci are still contributing to a very large fraction of the polygenetic variance component. The second explanation is that the loci are interacting in such a way that the variance contribution by multiple loci is less than the sum of the variance contributions of individual loci. In preliminary analyses of 188 unrelated individuals measured for the *apo E* polymorphism, *apo A-IV* polymorphism, and plasma levels of cholesterol, betalipoprotein and triglycerides, we have detected no significant evidence for statistical interaction between the effects of these loci on the levels of these quantitative phenotypes (Boerwinkle & Sing, unpublished results). This does not preclude interaction between other loci. The third explanation is that the loci with moderate and large effects have all been accounted for and the remaining loci have very small contributions to the phenotypic variance. The fourth explanation is that our methods of estimating the variance contribution of a single measured locus and the polygenetic variance component are incorrect and/or not comparable. The variance contribution of the single measured locus is calculated from its genotype frequencies and effects. The polygenetic variance component is a random effect estimated as a function of the covariance between relatives. The polygenetic model assumes there are a large number of unlinked loci, each with small effects so that, by the central limit theorem, the polygenetic effects are normally distributed in the population with variance σ_G^2 . The variance contribution by loci such as *apo E* is relatively large. How loci with large (fixed)

effects contribute to the polygenetic (random) effects is not clear. By the very nature and assumptions of the traditional polygenetic model, it may 'not allow' one to partition out of the estimated polygenetic variance component the effects of individual loci with large contributions to the variance of the quantitative phenotype.

The second major point is that the majority of the genetic variability can be accounted for by polymorphic gene loci with moderate effects on the quantitative phenotype. The contribution to cholesterol variability by genes marked by *ABO* and other marker loci (Sing & Orr, 1976) is small because of their minuscule effects. Loci with very large effects on the individual, such as *familial hypercholesterolemia* and its effect on cholesterol levels (Goldstein & Brown, 1979), for example, do not contribute greatly to the phenotypic variance in the population because of the rare frequency of the variant genotype. It is relevant for a discussion of the genetic architecture of a quantitative phenotype to distinguish between the relationship that exists between the effect of a single locus on the level of the phenotype for an *individual* and the effect of the locus on the variance of the phenotype in the *population* at large. We hypothesize that polymorphic loci in the population that have moderate effects on the level of the phenotype of an individual contribute to the majority of the genetic variance in the population.

This work was supported in part by NCI CA 26803 and NIH RO1 HL24489. We would like to thank Professor Gerard Siest and his colleagues at the University of Nancy, France for their support of this research. We would also like to thank Dr Samir Hanash and Dorothy Welsh for their help running the two-dimensional gels.

REFERENCES

- ALLAIN, C. C., POON, L. S., CHAN, C. S. G., RICHMOND, W. & FU, P. C. (1974). Enzymatic determination of total serum cholesterol. *Clin. Chem.* **20**, 470–475.
- BOERWINKLE, E. (1985). The use of measured genotype information in the genetic analysis of quantitative phenotypes. Ph.D. Thesis, University of Michigan, Ann Arbor.
- BOERWINKLE, E. & SING, C. F. (1986). Bias of the variance contribution of single loci. *Am. J. Hum. Genet.* **39**, 137–144.
- BOERWINKLE, E., CHAKRABORTY, R. & SING, C. F. (1986a). The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann. Hum. Genet.* **50**, 181–194.
- BOERWINKLE, E., VISVIKIS, S., WELSH, D., STEINMETZ, J., HANASH, S. & SING, C. F. (1986b). The use of measured genotype information in the analysis of quantitative phenotypes in man. II. The role of the apolipoprotein E polymorphism in determining levels, variability and covariability of cholesterol, betalipoprotein and triglycerides in a sample of unrelated individuals. *Am. J. Med. Genet.* (In the Press).
- BOERWINKLE, E., VISVIKIS, S., STRAHLER, J., STEINMETZ, J., HANASH, S. & SING, C. F. (1987). The use of measured genotype information in the analysis of quantitative phenotypes. IV. Frequency of the Apolipoprotein A-IV polymorphism and its effect on total cholesterol, betalipoprotein and triglyceride levels. *Am. J. Hum. Genet.* (Submitted).
- BOUTHILLIER, D., SING, C. F. & DAVIGNON, J. (1983). Apolipoprotein E phenotyping with a single gel method; application to the study of informative matings. *J. Lipid. Res.* **24**, 1060–1069.
- BRESLOW, J. L. (1985). Human apolipoprotein molecular biology and genetic variation. *Ann. Rev. Biochem.* **54**, 699–727.
- BURSTEIN, M. & SAMAILLE, L. (1959). Nouvelle méthode de séparation et de dosage des lipoprotéines de faible densité. *Ann. Biol. Clin.* **17**, 23–34.
- DAVIGNON, J., DUFOUR, R. & CANTIN, M. (1983). Atherosclerosis and hypertension. In *Hypertension* (eds J. Genest and O. Kuchel), pp. 810–852. New York: McGraw-Hill.
- GOLDSTEIN, J. L. & BROWN, S. (1979). The LDL receptor locus and the genetics of familial hypercholesterolemia. *Ann. Rev. Genet.* **13**, 259–290.
- HOPKINSON, D. A., SPENCER, N. & HARRIS, H. (1964). Genetical studies on human red cell acid phosphatase. *Am. J. Hum. Genet.* **16**, 141–154.
- HOUOT, O., GALTEAU, M. M., HENNY, G., HERBETH, B., SCHIELE, F., SIEST, G. & STEINMETZ, J. (1985). Laboratory organization and quality control. In *Interpretation of Clinical Laboratory Tests* (eds G. Siest, J. Henny, F. Schiele and D. Young), pp. 40–53. California: Biomedical Publications.

- MENZEL, H. J., KLADETSKY, R. G. & ASSMAN, G. (1983). Apolipoprotein E polymorphism and coronary artery disease. *Arteriosclerosis* **3**, 310–322.
- MOLL, P. P., KOTTKE, B. A., WEIDMAN, W., MICHELS, V. & SING, C. F. (1986). Genetic determination of plasma apo A-I levels: The Rochester family heart study. *Am. J. Hum. Genet.* (In the Press).
- NEEL, J. V., ROSENBLUM, B. B., SING, C. F., SKOLNICK, M. M., HANASH, S. M. & STERNBERG, S. (1984). Adapting two-dimensional gel electrophoresis to the study of the human germ line mutation rates. In *Methods and Applications of Two-Dimensional Gel Electrophoresis of Proteins* (eds. J. E. Celis and R. Bravo), pp. 259–306. New York: Academic Press.
- RAO, D. C., MORTON, N. E., GULBRANDSEN, C. L., RHOADS, G. G., KAGEN, A. & YEE, S. (1979). Cultural and biological determinants of lipoprotein concentrations. *Ann. Hum. Genet.* **42**, 467–477.
- ROBERTSON, F. W. & CUMMING, A. M. (1985). Effects of apoprotein E polymorphism on serum lipoprotein concentration. *Arteriosclerosis* **5**, 283–292.
- SAMMONS, D. W., ADAMS, L. D. & NISHIZAWA, E. E. (1981). A silver based color development system for staining of polypeptides in polyacrylamide gels. *Electrophoresis* **2**, 135–141.
- SING, C. F., CHAMBERLAIN, M., BLOCK, W. D. & FEILER, S. (1975). Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. I. Analysis of the frequency distribution for evidence of a genetic polymorphism. *Am. J. Hum. Genet.* **27**, 333–347.
- SING, C. F. & ORR, J. D. (1976). Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. III. Identification of genetic effects using 12 polymorphic genetic blood marker systems. *Am. J. Hum. Genet.* **28**, 453–464.
- SING, C. F. & ORR, J. D. (1978). Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. IV. Separation of polygene from common environmental effects. *Am. J. Hum. Genet.* **30**, 491–504.
- SING, C. F. & DAVIGNON, J. (1985). The role of Apolipoprotein E genetic polymorphism in determining normal plasma lipid and lipoprotein variation. *Am. J. Hum. Genet.* **37**, 268–285.
- STEINMETZ, J. & PANEK, E. (1978). Adaptation sur GSA II Greiner du dosage des triglycérides par voie entièrement enzymatique. Application à l'étude de leur conservation et de certaines interférences analytiques. *J. Clin. Chem. Clin. Biochem.* **16**, 613–619.
- TALMUD, P. J. & HUMPHRIES, S. (1986). DNA polymorphisms and the apolipoprotein B gene. *Lancet* *ii*, 104.
- THOMPSON, J. N., JR. (1975). Quantitative variation and gene number. *Nature* **258**, 665–668.
- THOMPSON, J. N. & THODAY, J. M. (1979). Synthesis; Polygenic variation in perspective. In *Quantitative Genetic Variation* (eds. J. N. Thompson and J. M. Thoday), pp. 295–302. London: Academic Press.
- UTERMANN, G., HEES, M. & STEINMETZ, A. (1977). Polymorphism of apolipoprotein E and occurrence of dysbetalipoproteinemia in man. *Nature* **269**, 604–607.
- UTERMANN, G., PRUN, N. & STEINMETZ, A. (1979). Polymorphism of apolipoprotein E. III. Effect of a single polymorphic gene locus on plasma lipid levels in man. *Clin. Genet.* **15**, 63–72.
- VOGEL, F. & MOTULSKY, A. G. (1979) *Human Genetics: Problems and Approaches*. New York: Springer-Verlag.
- ZANNIS, V. I. & BRESLOW, J. L. (1981). Human very low density lipoprotein: Apolipoprotein E isoprotein polymorphism is explained by genetic variation and post-translational modification. *Biochem.* **20**, 1033–1041.

APPENDIX

Hypothesis Testing for the Average Effects of Alleles Among Groups

The average effect of an allele may be interpreted as the expected deviation from the population mean of an individual carrying that allele. Most textbooks of quantitative genetics present expressions for the average effects of alleles as a function of the genotype frequencies and phenotype value (see for example, Kempthorne, 1969; Falconer, 1981). In this paper, we calculate the average effect of the three *apo E* alleles on total plasma cholesterol, betalipoprotein and triglyceride levels. The formula used here for estimating the average effect of an allele is equal to

$$\alpha_j = \frac{f_{jj}\mu_{jj} + \frac{1}{2}\sum_{i \neq j} f_{ji}\mu_{ji}}{f_j} - \mu;$$

where α_j is the average effect of the j th allele, i are the other alleles at the locus, μ_{ij} is the mean phenotypic level of the ij th genotype, μ is the grand mean, f_{ij} is the expected frequency of the ij th genotype assuming Hardy–Weinberg equilibrium and f_j is the allele frequency of the j th

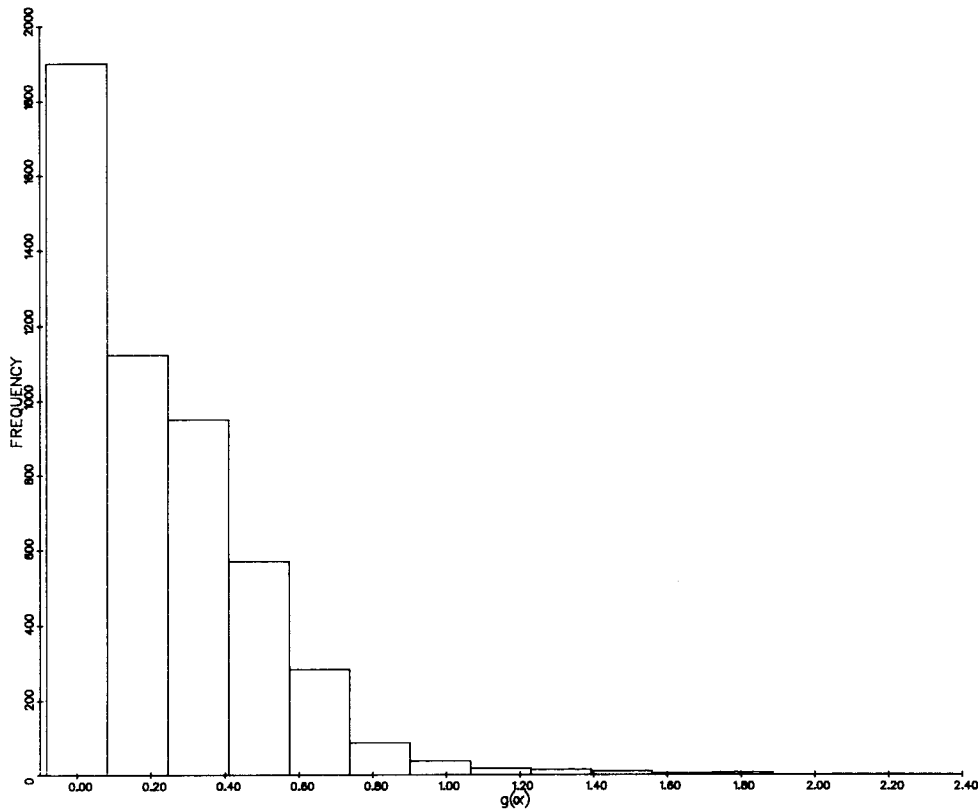


Fig. 2. Frequency distribution of $g(\alpha)$ for total cholesterol levels over 5000 permutations of the data.

allele. For a given genetic system there are as many average effects of alleles as there are alleles at that locus. The average effects of alleles at a locus are subject to the restriction: $\sum_j f_j \alpha_j = 0$. The statistical properties of the estimated average effects are not known. They are complicated functions of the estimated allele frequencies and genotype means. A method for testing hypotheses about the value of the average effects of alleles at a locus, or the equality of the average effects of the alleles at a locus among groups is not obvious.

We propose and have applied a sampling-resampling approach to test the equality of a vector of average effects of alleles among groups. This approach is an extension of an approach suggested by Templeton *et al.* (1987) to test whether the average effects of the alleles at a locus are statistically significantly different from zero. The statistic we propose is:

$$g(\alpha) = \sum_k \sum_j (\alpha_{jk} - \alpha_j)^2,$$

where α_{jk} is the average effect of the j th allele in the k th group and α_j is the average effect of the j th allele ignoring the groups. It is understood that the summation over the j alleles precedes the summation over the k groups. The distribution of this statistic is not known. The null distribution of the statistic may be approximated by calculating the value of the statistic from repeated sampling data under the null hypothesis (Efron, 1982). These repeated samples are generated from the sample of data to be tested by resampling without replacement from the data and allocating the observations randomly to the k groups. The overall genotype-phenotype relationship (i.e. the average effects of the alleles ignoring groups = α_j) and the size

of the k groups is preserved in these analyses. If the value of $g(\alpha)$ from the data exceeds the percentile of the null distribution of $g(\alpha)$ determined by the desired significance level, one rejects the null hypothesis that the average effects of the alleles at this locus are the same among the k groups.

For the present application, we wished to determine whether the average effects of the *apo E* alleles are homogeneous among the parents and the youngest child from each family. We carried out 5000 permutations of the data among groups before any inference was made. We determined that 5000 permutations were sufficient for the distribution of $g(\alpha)$ to stabilize. For total cholesterol levels, the value of $g(\alpha)$ for the data is 0.091. The sample frequency distribution of $g(\alpha)$ for the 5000 permutations of the data under the null hypothesis is given in Figure 2. The cut-off values that determine the 10th, 5th and 1st percentile of this distribution are 0.537, 0.647, 0.865, respectively. The statistic $g(\alpha)$ for the data does not exceed any of these significance levels. Therefore, we infer that the average effects of the *apo E* alleles on total cholesterol levels among the parents and the youngest child from each family are homogeneous. The average effects of the *apo E* alleles on betalipoprotein and triglyceride levels were also not statistically significantly different between these two groups.

REFERENCES

- EFRON, B. (1982). The jackknife, the bootstrap and other resampling plans. Society for Industrial and Appl. Mathematics. Philadelphia, PA.
- FALCONER, D. S. (1981). *Introduction to Quantitative Genetics*. New York: Longman.
- KEMPTHORNE, O. (1969). *Introduction to Genetic Statistics*. Iowa: Iowa State University Press.
- TEMPLETON, A., SING, C. F., KESSLING, A. & HUMPHRIES, S. (1987). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. II. Associations of triglycerides and cholesterol levels with apolipoprotein A-I, C-III, A-IV region in a human population. (Submitted).