

An Application of a Model for a Genotype-Dependent Relationship Between a Concomitant (Age) and a Quantitative Trait (LDL Cholesterol) in Pedigree Data

Patricia P. Moll, Charles F. Sing, Suzanne Lussier-Cacan, and Jean Davignon

Departments of Epidemiology (P.P.M.), and Human Genetics (P.P.M., C.F.S.), University of Michigan, Ann Arbor; Department of Lipid Metabolism and Atherosclerosis Research, Clinical Research Institute of Montreal (S.L.-C., J.D.)

In most genetic studies in humans the variability in a quantitative trait is adjusted for variability in concomitants (age, sex, etc) using a single regression equation prior to analyses of pedigree data. To illustrate an alternative approach, a single locus genetic model was tested. This model incorporates genotypic effects on the level of the trait, the variability in the trait, and the relationship between a concomitant and the trait. In this study, the model was applied to measures of age and low-density lipoprotein (LDL) cholesterol in a large kindred with familial hypercholesterolemia. The application of this model to 322 individuals in four generations provided evidence that genotypic variation at a single locus influences LDL levels early in life, the rate of increase of LDL with age and the phenotypic variance. A model with genotype-dependent slope and variance fit the data significantly better than a model with slope and variance independent of genotype. The inclusion of age-specific genotypic differences contributed to identification of high-risk individuals, to statistical support for a major locus, and to evidence for genetic determination of the tracking of LDL levels. Models that incorporate genotype-specific concomitant effects have the potential to represent more realistically the relationship between genotypic variability and quantitative phenotypic variation than models that assume that these effects do not exist.

Key words: segregation analysis, familial hypercholesterolemia, concomitants, quantitative genetics, low-density lipoprotein cholesterol

INTRODUCTION

Preliminary evidence for segregation at a genetic locus with a major effect on the quantitative variability of a trait is often obtained by fitting a mixture of normal

Received for publication May 16, 1984; revision received July 24, 1984.

distributions to the data (Bucher et al, 1982; Burns, 1982; McGue et al, 1983; Boerwinkle et al, 1984). If a mixture of two or three normal distributions fits significantly better than a single normal distribution when the data are considered as derived from a sample of unrelated individuals, then segregation analysis is warranted [Elston et al, 1975]. There are several approaches to modeling the segregation of a major locus effect in pedigrees [Elston and Stewart, 1971; Morton and McLean, 1974; Cannings et al, 1978; Lalouel et al, 1983], and each has been applied to study the transmission of various quantitative traits [Elston et al, 1975; Namboodiri et al, 1977; Morton et al, 1978; Williams and Lalouel, 1982; Darlu et al, 1983; Hasstedt et al, 1983a; Boerwinkle et al, 1984; Moll et al, 1984]. Prior to the genetic analysis in each of these studies, standard regression methods were used to adjust variability in the quantitative trait for variability in one or more concomitants such as age and sex. This adjustment approach assumes that the relationship between concomitants and the quantitative variable is the same for each of the major locus genotypes.

Murphy [1979a] presented the expected distribution of a quantitative trait when there is a major-locus genotype-specific relationship with age. Although Murphy did not consider the statistical properties of an age-dependent genetic effect, he concluded that the best evidence for a mixture of distributions and the possible existence of a major locus effect would come from investigations of the trend toward bimodality with increasing age.

Here we present an application of segregation analysis that incorporates single-locus genotype-specific effects on the relationship between age and the low-density lipoprotein cholesterol (LDL) levels in a large French-Canadian kindred with familial hypercholesterolemia, a known mendelian disorder [McKusick, 1983]. Our study establishes that, in addition to a general effect on LDL levels, the major locus genotype also affects the rate of increase of LDL with age and the phenotypic variance. We found that incorporation of genotype-specific parameters for these differences into a genetic model improves the identification of individuals at high risk for having the mutant allele and increases the support for a single major-locus effect on LDL cholesterol variability. This example establishes that models which include a genotype-specific relationship between concomitants and quantitative traits have the potential to increase our understanding of the relationship between genotype and phenotype.

FAMILIAL HYPERCHOLESTEROLEMIA

Familial hypercholesterolemia (FH) is characterized by a very high LDL level, the presence of tendon xanthomas in many of the affected subjects, and premature atherosclerosis [Goldstein and Brown, 1983]. This condition is an autosomally inherited trait. The prevalence of heterozygotes in the United States is estimated to be between 1 in 500 and 1 in 1,000 [Vogel and Motulsky, 1979]. The genetic locus determining FH is on chromosome 19 [McKusick, 1984]. Heterozygotes have moderate hypercholesterolemia [350 to 550 mg/dl] from birth, while the rare homozygotes (estimated to be about 1 in a million persons) have severe hypercholesterolemia (650 to 1,000 mg/dl). Serum cholesterol is carried mostly by LDL. The mean LDL concentration in heterozygotes is approximately two to three times the mean for normal individuals. For homozygotes, the mean LDL level is approximately six times that for normals. Homozygotes with FH will have coronary heart disease (CHD) in

their teens. Approximately half of the male heterozygotes and a fifth of the female heterozygotes will have overt clinical manifestations of CHD by age 50 [Goldstein and Brown, 1983]. The primary genetic defect in FH results from one of several mutations in the gene specifying the extrahepatic cell surface receptor that controls the degradation of plasma LDL [Goldstein and Brown, 1983].

MATERIALS AND METHODS

The Kindred

A study was initiated to investigate the etiology of hypertriglyceridemia in individuals with FH type IIb. The index case that identified the kindred studied here was selected from a sample of 26 unrelated probands with type IIb hyperlipidemia and well-documented FH who were seen at the Clinical Research Institute of Montreal. The criteria for selection included the number of living relatives and their geographic dispersion. The index case was a 36-year-old man with bilateral xanthelasma, arcus corneae, and tendon xanthomas. His fasting plasma cholesterol level was 370 mg/dl; plasma LDL, 261 mg/dl; and plasma triglycerides, 238 mg/dl. He had the expected partial deficit of LDL-receptors, as measured in surviving lymphocytes by ^{125}I -LDL degradation. Shortly before initiation of collection of lipid and lipoprotein measurements in his relatives, he sustained (at age 37) an acute antero-septal myocardial infarction from which he recovered.

Of the 508 individuals identified as belonging to the kindred identified by this index case, 378 (four generations) were studied. These 378 individuals included all living first- and second-degree relatives of the index and over 75% of the living, more genetically distant relatives. Of the 378 studied, 321 ranging in age from 2 to 83 years had lipoprotein levels measured. A detailed description of the index case, the individuals in the kindred, the sampling procedures, and laboratory methods is presented elsewhere [Davignon et al, 1983]. Any individual in the kindred who presented with one or more influences known to affect lipid levels (nonfasting, drugs, diet, current pregnancy, recent illness, etc) was evaluated to determine whether resampling was necessary. The distribution of plasma lipid and the lipoprotein abnormalities in the relatives of the index case were consistent with the presence of two separate hereditary lipid disorders: FH in the paternal relatives and familial hyperpre-beta lipoproteinemia in the maternal relatives [Davignon et al, 1983]. No one studied in this kindred was homozygous for FH.

Least squares regression analyses [Neter and Wasserman, 1974], which assume that every observation in the kindred is uncorrelated with every other measurement, were used to establish the presence of a relationship between age and LDL in this kindred. There was a significant linear (but not quadratic nor cubic) relationship ($p < 0.05$) that was not significantly different between males and females. Hence, the genotype-dependent relationship will be restricted to that of age and LDL. The overall estimates of the intercept and slope were 112.39 and 1.06, respectively. Variation in age explained approximately 7.7% of the variation for LDL.

A previous study of this kindred reported heterogeneity in the age-LDL relationship among groups of individuals identified by their relationship to the index case [Davignon et al, 1980]. Further evidence for heterogeneity of the age-LDL relationship was found between two subsets having different clinically defined lipid phenotypes. Individuals were previously assigned to a specific lipid phenotype [Davignon

et al, 1983] following the criteria of Fredrickson et al [1967] with the additional separation of type IIa from type IIb according to the triglycerides-to-total cholesterol ratio, with type IIb having a total triglycerides-to-total cholesterol ratio of 0.40 or greater. Sixteen individuals could not be classified because of an inconsistency in the interpretation of the lipid and lipoprotein measurements. There were no type I or type III individuals in the kindred. The 48 individuals considered to be potential carriers of a mutant allele for FH (type IIa and type IIb individuals related to the index case through his father) were contrasted to the 274 other individuals for the age-LDL relationship with the use of least squares techniques. Between these two groups, the hypothesis of equal slopes was rejected ($p < 0.01$). The estimates of the intercepts and slopes were 208.9 and 1.77 for the individuals considered to be potential carriers of the mutant allele and 95.1 and 0.95 for the remaining individuals in the kindred. Variation in age explained approximately 38% of the variation in LDL in the former group and approximately 24% in the latter.

A Genotype-Specific Age-Regression Model

We begin by reviewing the model for the segregation of a major locus effect proposed by Elston and Stewart [1971]. In the application to the kindred considered here, we assume that one autosomal genetic locus has a major effect on the quantitative phenotype. Variation in the LDL cholesterol among individuals with the same major-locus genotype, which we shall write as g , is assumed to result solely from environmental effects specific to each individual. This variation is denoted $\sigma_{e|g}^2$ ($g = aa, Aa, \text{ or } AA$). We assume that in the population sampled there is neither gene-environment correlation nor environmental correlation between individuals within the pedigree and that Hardy-Weinberg equilibrium and random mating exist in the population from which the kindred was sampled. The major-locus model is usually formulated under the assumption of independence of the genetic and environmental effects. We first review this model and then turn to a presentation of modifications to include genotype-dependent age regression and genotype-dependent environmental variance.

We assume that the locus with the major effect on LDL has two alleles, a and A , and that the relative frequency of allele a is p . The probability, $f(g)$, that an individual chosen at random from the population will have genotype g is predicted by the Hardy-Weinberg law, so that $f(aa) = p^2$, $f(Aa) = 2(1-p)p$ and $f(AA) = (1-p)^2$. The genotype distribution of offspring is conditioned on the genotypes of their parents. In general, $t(g|g_f, g_m)$ is the probability that a child has genotype g , given that the genotypes of the father and mother are g_f and g_m , respectively. Mendelian segregation is assumed. A penetrance function, $p(x|g)$, defines the probability-density function of the quantitative phenotype, x , conditional on the g th genotype. If this distribution is assumed to be normal with mean μ_g and variance $\sigma_{e|g}^2$, the expression for the penetrance is

$$p(x|g) = [1/(2\pi\sigma_{e|g}^2)^{1/2}] \exp[-(x - \mu_g)^2/2\sigma_{e|g}^2]. \quad (1)$$

Data are related to the parameters associated with allele frequency, transmission of genes, and penetrance by a likelihood function, L , which summarizes all the information available from the measured phenotypes regarding the unknown model

parameters. The likelihood of a pedigree of n individuals may be written

$$L = \prod_{g_1} \prod_{g_2} \dots \prod_{g_n} \prod_i p(x_i | g_i) \prod_j f(g_j) \prod_k t(g_k | g_{k_m}, g_{k_f}), \tag{2}$$

where the first product is over each individual i in the kindred whose phenotype was observed (i = 1, . . . ,n), the second product is over each original j (the individuals whose parents are not included in the kindred), and the third product is over each non-original k (individuals who have at least one ancestor included in the study). Each original contributes a term of the form $\sum f(g)p(x | g)$ to the likelihood while each non-original contributes a term of the form $\sum t(g | g_f, g_m)p(x | g)$, where in each case the sum is over all possible genotypes for that individual. Equation 2 is, by definition, the likelihood of observing all the phenotypes in the pedigree conditional on the pedigree structure and on the parameters that specify $p(x | g)$, $f(g)$, and $t(g | g_f, g_m)$.

The relationship between genotype and phenotype modeled by Equation 1 does not allow for inclusion of a genotype-dependent relationship between the quantitative trait, x, and a concomitant variable. A more general form of the penetrance function may be written

$$p(x | g) = [1/(2\pi\sigma_{e|g}^2)^{1/2}] \exp \{ -[x - (\alpha_g + \beta_g \cdot \text{age})]^2 / 2\sigma_{e|g}^2 \} \tag{3}$$

where the intercept and slope of the age regression, and the individual-specific genotype-dependent environmental variance for the gth genotype (g = aa,Aa,AA) are α_g , β_g and $\sigma_{e|g}^2$, respectively.

Eight different models (Table I) can be defined which reflect the presence or absence of a genetic influence on these three parameters. Model 8 assumes that the three parameters are equal for the different genotypes—ie, that there is a single normal distribution for LDL at every age that is independent of genotype at the major locus. Because differences in LDL levels are known to exist at birth between individuals with and without the mutant allele for FH, we did not consider models 5, 6 and 7, which assume equal intercepts. Only dominant models were considered in this application since it is agreed that the rare allele (A) that elevates LDL is dominant

Table I. Eight Different Models Based on the Presence (Yes) or Absence (No) of a Genetic Influence on the Intercept and Slope of the Age Regression and the Individual Specific Environmental Variance

Model	Intercept	Slope	Environmental variance
1	Yes	Yes	Yes
2	Yes	Yes	No
3	Yes	No	Yes
4	Yes	No	No
5	No	Yes	Yes
6	No	Yes	No
7	No	No	Yes
8	No	No	No

over the more common allele (a) that is associated with normocholesterolemia [Goldstein and Brown, 1983; McKusick, 1983]. Model 4 assumes that only the intercept is genotype dependent ($\alpha_{aa} < \alpha_{Aa} = \alpha_{AA}$, assuming dominance). This model has been used by others [Elston et al, 1974]. The role of genotype in model 4 corresponds to the role of genotype assumed by first adjusting the phenotype for age variability and then fitting genetic models. Model 3 assumes both the intercept and variance are genotype dependent ($\alpha_{aa} < \alpha_{Aa} = \alpha_{AA}$, $\sigma_{e|aa}^2 \neq \sigma_{e|Aa}^2 = \sigma_{e|AA}^2$). Model 2 assumes that the slope and intercept are genotype dependent ($\alpha_{aa} < \alpha_{Aa} = \alpha_{AA}$, $\beta_{aa} \neq \beta_{Aa} = \beta_{AA}$), while model 1, the full model among those considered, assumes all three parameters to be genotype dependent ($\alpha_{aa} < \alpha_{Aa} = \alpha_{AA}$, $\beta_{aa} \neq \beta_{Aa} = \beta_{AA}$, $\sigma_{e|aa}^2 \neq \sigma_{e|Aa}^2 = \sigma_{e|AA}^2$).

Comparisons of the likelihoods and maximum likelihood estimates (MLE) of parameter values in models 2, 3, and 4 to model 1 test the effects of different genotypes on the dispersion of phenotypic values (ie, the $\sigma_{e|g}^2$) and/or relationship between age and phenotype (ie, the β_g), given that genotype affects phenotype displacement [ie, $\alpha(aa) < \alpha(Aa) = \alpha(AA)$]. Model 8 represents the null hypothesis that the distribution of LDL given age is independent of genotype at a single locus with a major effect in this kindred.

Comparisons between two models can be made by computing the ratio of the maximized likelihood of a full model to the maximized likelihood of a reduced model. The natural logarithm of this ratio multiplied by negative 2 is distributed approximately as a chi-square if certain assumptions hold. Although some of the assumptions may not hold in this application of likelihood analysis to a single pedigree [Cannings et al, 1980], chi-square statistics are presented along with common logarithms of this ratio. The likelihoods of the models considered were computed using the Pedigree Analysis Package [Hasstedt and Cartwright, 1979; Hasstedt et al, 1979] and maximized with the use of a quasi-Newton method [Lalouel, 1979]. To correct for ascertainment, the likelihood of the pedigree was divided by the likelihood of observing the phenotype of the index case [Cannings and Thompson, 1977].

The likelihood, L , in Equation 2, contains one summation over the three possible genotypes at the major locus for each individual in a pedigree. This likelihood function may be used to compute the probability of each genotype for each individual in the pedigree conditional on the pedigree structure, the phenotypic values for the individuals in the pedigree, the model selected, and the MLEs of the parameters of the model. The conditional likelihood, $L(g_i)$, that a specific individual, i , has a specific genotype, g_i , is obtained by replacing that individual's summation in the likelihood for the pedigree by a single term corresponding to a specific g_i . The probability that a designated individual has the genotype g_i , conditional on what is known about the pedigree, is equal to $L(g_i)/\Sigma L(g_i)$, where the summation is over the three major locus genotypes for individual i .

RESULTS

The maximum likelihood estimates of parameters and their standard errors under models 1-4 and model 8 are presented in Table II. The \log_{10} likelihoods for each model are compared to the \log_{10} likelihood for model 1. The maximum likelihood estimates of the intercept, slope, and variance under model 8, as expected, were in close agreement with those obtained in the preliminary analysis of lipid phenotypes

TABLE II. Maximum Likelihood Estimates of Parameters

Parameter	Model 1	Model 2	Model 3	Model 4	Model 8
	$\alpha_{aa} < \alpha_{Aa} = \alpha_{AA}$ $\beta_{aa} \neq \beta_{Aa} = \beta_{AA}$ $\sigma_{e aa}^2 \neq \sigma_{e Aa}^2 = \sigma_{e AA}^2$	$\alpha_{aa} < \alpha_{Aa} = \alpha_{AA}$ $\beta_{aa} \neq \beta_{Aa} = \beta_{AA}$ $\sigma_{e aa}^2 = \sigma_{e Aa}^2 = \sigma_{e AA}^2$	$\alpha_{aa} < \alpha_{Aa} = \alpha_{AA}$ $\beta_{aa} = \beta_{Aa} = \beta_{AA}$ $\sigma_{e aa}^2 \neq \sigma_{e Aa}^2 = \sigma_{e AA}^2$	$\alpha_{aa} < \alpha_{Aa} = \alpha_{AA}$ $\beta_{aa} = \beta_{Aa} = \beta_{AA}$ $\sigma_{e aa}^2 = \sigma_{e Aa}^2 = \sigma_{e AA}^2$	$\alpha_{aa} = \alpha_{Aa} = \alpha_{AA}$ $\beta_{aa} = \beta_{Aa} = \beta_{AA}$ $\sigma_{e aa}^2 = \sigma_{e Aa}^2 = \sigma_{e AA}^2$
α_{aa}	94.31 ± 2.8	94.78 ± 3.1	95.21 ± 2.7	92.30 ± 3.0	112.37 ± 5.7
$\alpha_{Aa} = \alpha_{AA}$	203.21 ± 10.8	207.12 ± 7.8	223.17 ± 6.4	225.13 ± 5.0	112.37 ± 5.7
β_{aa}	0.94 ± 0.1	0.95 ± 0.1	0.99 ± 0.1	1.06 ± 0.1	1.05 ± 0.2
$\beta_{Aa} = \beta_{AA}$	1.83 ± 0.4	1.80 ± 0.3	0.99 ± 0.1	1.06 ± 0.1	1.05 ± 0.2
$\sigma_{e aa}^2$	583.71 ± 27.8	731.70 ± 31.9	581.77 ± 27.7	748.02 ± 32.8	2983.69 ± 118.5
$\sigma_{e Aa}^2 = \sigma_{e AA}^2$	1428.84 ± 162.16	731.70 ± 31.9	1559.46 ± 178.1	748.02 ± 32.8	2983.69 ± 118.5
P	0.987 ± .009	0.995 ± .007	0.986 ± .010	0.994 ± .008	
$\log_{10}(L) - \log_{10}(Lr)^a$	—	3.3866	1.1128	5.1821	66.2500
$2[\ln(L) - \ln(Lr)]^a$	—	15.5959	5.1246	23.8645	305.0925

^aL = likelihood of model 1; Lr = likelihood of a reduced model.

with the use of least squares techniques to establish the presence of an age regression. Since the approximate chi-square statistic for comparing model 8 and model 1 is 305, we reject the model of no genotype effects.

In comparing models 1-4, the estimates of allele frequencies are similar. The differences between genotype-specific intercepts are smaller for models with a genotype-dependent slope (models 1 and 2) than for models with equal slopes (models 3 and 4). In the models with genotype-dependent slopes, the rate of change with age for those individuals with a mutant allele is approximately twice that for individuals with the normal aa genotype. In models with genotype-dependent individual-specific environmental variance (models 1 and 3), the variance for those individuals with the mutant allele is approximately 2.5 times that for individuals with the normal aa genotype, regardless of whether the slope is genotype-dependent or not. The approximate chi-square statistics in Table II argue strongly against models 2, 3, and 4 in favor of model 1.

Approximate asymptotic correlations were computed from the estimated variance-covariance matrix at the maximum likelihood estimates to assess the interdependence of the parameter estimates. These correlations are given in Table III for model 1 and model 8. In these models, as well as in the others considered, there is a strong negative correlation between estimates of slope and intercept. This is an expected result under a linear regression [Neter and Wasserman, 1974]. The estimates for allele frequency and for individual-specific environmental variance are not strongly correlated with the estimates for other parameters in the models. Based on the approximate chi-square test of reduced models, on the confidence intervals that can be constructed for the estimates of the parameters in model 1, and on the lack of correlation between the MLEs for $\sigma_{e|g}^2$ and the other parameters, it appears that not only are the differences early in life (the intercepts) genotype dependent but also the increase in LDL with increasing age (the slopes) and individual-specific environmental variance are different between genotypes.

To identify individuals at high risk for coronary heart disease, we estimated the probability of carrying the mutant allele (Pr(Aa)) for each individual in the kindred.

Table III. Correlations Among Maximum Likelihood Estimates of Parameters

		Model 1						
	P	α_{aa}	$\alpha_{Aa} = \alpha_{AA}$	β_{aa}	$\beta_{Aa} = \beta_{AA}$	$\sigma_{e aa}^2$	$\sigma_{e Aa}^2 = \sigma_{e AA}^2$	
P	1.000							
α_{aa}	-0.001	1.000						
$\alpha_{Aa} = \alpha_{AA}$	0.204	0.041	1.000					
β_{aa}	0.055	-0.841	-0.026	1.000				
$\beta_{Aa} = \beta_{AA}$	0.040	-0.031	-0.850	0.033	1.000			
$\sigma_{e aa}^2$	0.175	0.021	0.038	0.031	0.014	1.000		
$\sigma_{e Aa}^2 = \sigma_{e AA}^2$	-0.086	-0.028	-0.141	0.001	0.077	-0.069	1.000	
		Model 8						
		$\alpha_{aa} = \alpha_{Aa} = \alpha_{AA}$	$\beta_{aa} = \beta_{Aa} = \beta_{AA}$		$\sigma_{e aa}^2 = \sigma_{e Aa}^2 = \sigma_{e AA}^2$			
$\alpha_{aa} = \alpha_{Aa} = \alpha_{AA}$		1.000						
$\beta_{aa} = \beta_{Aa} = \beta_{AA}$		-0.840	1.000					
$\sigma_{e aa}^2 = \sigma_{e Aa}^2 = \sigma_{e AA}^2$		0.010	-0.010		1.000			

The full model (model 1) and a reduced model (model 4) evaluated at their MLEs were each used to obtain separate estimates of Pr(Aa). These two estimates were compared to assess the effect of excluding the genotype-specific effects on slope and variance on the estimates of Pr(Aa). For each of 314 of the 322 individuals in the kindred, the difference between the two estimates of Pr(Aa) was less than 0.10. Forty-five of these individuals had a high probability (> 0.90) of having the mutant allele under both models, while the remaining individuals had a low probability (< 0.10). The remaining eight individuals represent those for whom the estimate of Pr(Aa) was very different between models. The difference ranged from 0.10 to 0.57, with a mean of 0.33. Data on these eight individuals are presented in Table IV. All of these individuals were less than age 40, and none were diagnosed as having atherosclerosis, a myocardial infarction, or tendon xanthomas at the time the levels of LDL cholesterol were measured.

If we assume that the full model provides the most accurate estimate of the probability of carrying the mutant allele for all individuals in this kindred, then for seven individuals we would have underestimated Pr(Aa) and for one individual we would have overestimated Pr(Aa) if Pr(Aa) had been calculated under the reduced model rather than the full model. In this kindred with FH, where there is a large displacement between genotype means for LDL at every age, there is a group of rather young individuals for whom the additional information about age-specific genotype effects influences the assessment of probability of carrying the mutant allele.

The measured level of LDL plotted against age for the 322 individuals in this kindred is shown in Figure 1. Those individuals with Pr(Aa) greater than 0.90 under the full model are distinguished from those with Pr(Aa) less than 0.90. The genotype-dependent linear relationships (± 1 standard deviation) that have been predicted from the MLEs under the full model (model 1) are superimposed upon the data.

DISCUSSION

Models 1 through 4 considered here are special cases of a general model presented by Murphy [1979a] for the distribution of a quantitative trait in which the level is influenced both by genotype at a single locus and by age. The approach of incorporating the concomitant effects in the model rather than controlling for those effects prior to assessing the role of genotype on phenotype should provide additional

Table IV. Characteristics of Eight Individuals With Variable Estimates of Pr(Aa)

Age	Sex	LDL value	Triglyceride/cholesterol	Pr(Aa) full model ^a	Pr(Aa) reduced model ^b
34	M	204	0.35	0.12	0.01
7	M	155	0.38	0.68	0.12
13	F	299	0.08	0.74	0.93
39	M	223	1.62	0.78	0.21
12	F	170	0.23	0.88	0.44
12	M	174	0.14	0.93	0.61
33	M	231	0.32	0.98	0.64
28	F	215	0.35	0.98	0.88

^aModel 1 in text

^bModel 4 in text

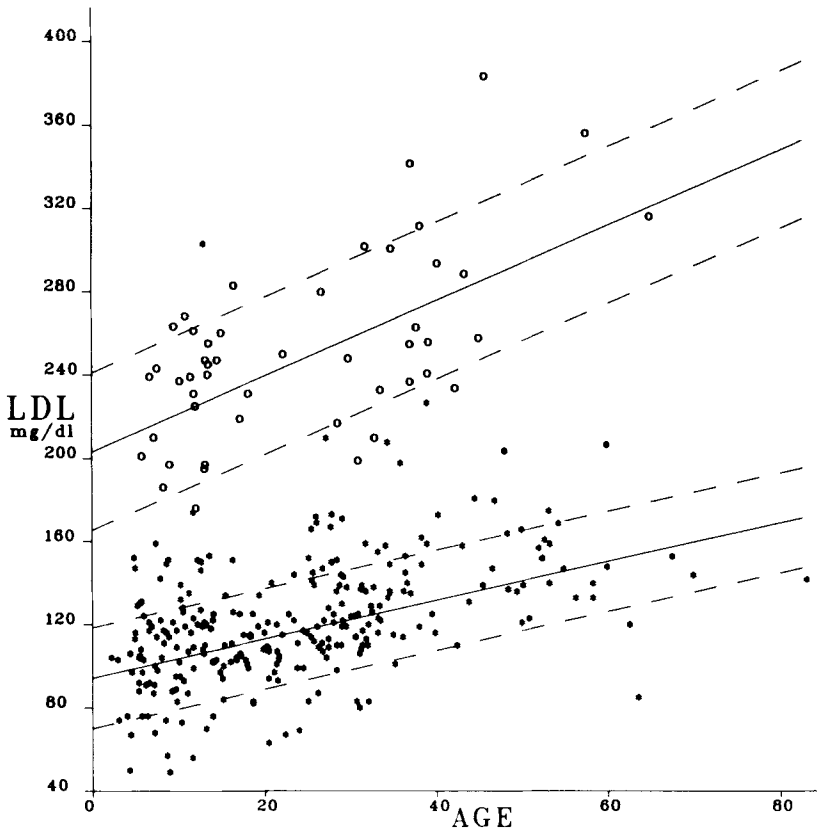


Fig. 1. The levels of LDL cholesterol are plotted against age for individuals (o) assumed to have Aa genotype ($\text{Pr}(Aa) > 0.90$) and individuals (*) assumed to have aa genotype ($\text{Pr}(Aa) < 0.90$). The predicted genotype-specific age-LDL relationships under model 1 for Aa and aa genotypes (± 1 standard deviation) are superimposed upon the data.

support for segregation at a single locus if the locus influences the change in the trait with increasing age (β) and/or the residual individual-specific variances (σ_e^2), as well as overall differences (α). Restricting the slopes and variances to be equal for all genotypes (model 4 vs model 1) reduced the \log_{10} likelihood by 5.18 and provided an approximate chi-square statistic of 23.86 ($p < 0.01$, $df = 2$) for this comparison in this kindred. The approximate chi-square statistics were 15.60 ($p < 0.01$, $df = 1$) and 5.12 ($p < 0.05$, $df = 1$) for model 2 (genotype-dependent slope) versus model 1 and model 3 (genotype-dependent variance) versus model 1, respectively. Others have used approximate chi-square statistics of similar magnitudes to reject a null hypothesis in genetic analysis [Namboodiri et al, 1977; Beaty et al, 1983; Hasstedt et al, 1983b].

Additional information about gene action beyond the effect on displacement between genotype means can be inferred from model 1. While the data in this kindred were cross-sectional, the results are consistent with the role of genetic factors in the determination of the tracking of LDL. Tracking is the persistence of an individual to remain at the same rank within a distribution of peers over time. The MLEs in model 1 predict that the individuals with lower initial values (α_{aa}) will have a lower mean value at any specific age ($\alpha_{aa} + \beta_{aa} \cdot \text{age}$) compared to the mean of those individuals

with higher initial values (the Aa individuals). Tracking for LDL has been reported recently for several human populations studied longitudinally. A study of a cohort of 440 children found that the levels of LDL at 6 months were significantly correlated with those at 1 year ($r = 0.50$) and 7 years ($r = 0.35$) [Webber et al, 1984]. Criqui et al [1983] reported that LDL values measured 6 years apart in 614 adults were significantly correlated ($r = 0.68$). A study of a cohort of aviators found a correlation of 0.63 for LDL for the 12 years from age 42 to age 54 (Harlan et al, 1984). The strongest predictor of change over the 12-year period among the aviators was the level at age 42 for LDL. The finding that those with the highest initial levels had the greatest change over the 12-year period is consistent with the estimate of β_{Aa} being larger than the estimate of β_{aa} in model 1. In addition to tracking in longitudinal data, there is often another phenomenon: that an individual with an extreme value on an early measurement will tend to be closer to the mean for a later measurement [Davis, 1976].

Cross-sectional data that are available are also consistent with our findings. The Lipid Research Clinics (LRC) Population Studies Data reported in 1980 (visit 2, random sample) for whites pooled for males and females suggest an average increase of approximately 8 mg/100 dl per decade for LDL for normocholesterolemics. For those with aa genotype in our kindred, model 1 predicts an increase of 9.4 mg/100 dl per decade. Our predicted levels for LDL for aa genotype are not significantly different from the means reported for LDL from the LRC.

A separate study of 220 males and 222 females presumed to be heterozygotes for FH provides a comparison with those in our kindred assumed to have the Aa genotype [Gagne et al, 1979]. For each decade, the predicted level of LDL for individuals with Aa genotype in our kindred did not vary more than one standard deviation from the means reported by Gagne et al [1979]. Our model 1 predicts an increase of LDL of 18.3 mg/100 dl per decade, while we estimate that the increase per decade up to age 50 was 19.3 mg/100 dl in the sample reported by Gagne et al [1979].

The factors that contribute to the increase in lipid and lipoprotein levels as age increases, regardless of genotype, are not well understood. Age is an index of a number of environmental factors that might be expected to change over the lifetime of the individual; eg, maturation, diet, physical activity, total body size, hormone levels, and stress may all influence the expression of genes that are involved in determining lipid and lipoprotein levels. Most populations of humans and many animal species show an increase of total cholesterol with age [Sabine, 1977]. The increase in LDL in humans is also well documented [Myant, 1981].

The genotype-specific estimates of variance in model 1 are consistent with variances in LDL levels reported in the LRC random sample and the study of heterozygotes for FH. Comparison of these two samples shows that at all ages, the heterozygotes for FH have a variance that is at least 2.5 times the variance for a random sample that is composed primarily of normocholesterolemic individuals.

Genotype assignment in this kindred represents an extreme situation in quantitative genetics. With such a small overlap between the distributions of LDL for those with and without the mutant allele, the misclassification of genotype should be minimized. We estimate that 87% of a random sample of unrelated individuals age 30 years would lie outside the region of overlap between the distributions of LDL for those with and without the mutant allele. In any sample enriched for

individuals with high LDL levels, misclassification will be increased. If model 1 does provide the most accurate probability of genotype, we would have underestimated the probability of having the mutant allele for seven individuals in this kindred and overestimated for one by using all the information in the kindred but ignoring the genotypic effect on variance and the relationship between age and LDL. Under model 1 we could identify 31 individuals in this kindred who were under age 40 and who had the mutant allele ($\text{Pr}(Aa) > 0.90$) but no atherosclerosis or tendon xanthomas. These individuals represent a group who are undoubtedly at high risk for developing disease. Three of these 31 would have been excluded from this high-risk group under model 4. For traits other than LDL in which the displacement between genotypic means is not as great, the inclusion of the genotype-specific variance and the genotypic-specific relationship between a concomitant and a quantitative trait, when these exist, might be expected to markedly improve the accuracy of predicting which individuals in the kindred are at high risk for having the deleterious allele.

An approach for analyzing pedigree data that is stepwise (first regression analysis and then fitting models of transmission) is computationally easier than the approach presented here. However, if genes are involved in the regression relationship, testing hypotheses about this role of genes is not possible in the stepwise approach. Murphy [1979b] has called to attention the preoccupation with estimation of first and second moments in quantitative genetic studies and the failure to consider more realistic models that incorporate higher-order interactions. The example presented here illustrates the consequences for modeling genotype-specific phenotypic variability and relationships with concomitants.

Many studies of quantitative traits begin by applying a transformation to the data. Converting the data into their logarithms is a common approach [Elston et al, 1974; Elston et al, 1975; Namboodiri et al, 1977; Beaty et al, 1983; Hasstedt et al, 1983b; Moll et al, 1984], since distributions skewed to the right are often made more symmetric by transformation to logarithmic scale and such a transformation is likely to make the variance less dependent on the mean [Sokal and Rohlf, 1969]. We applied a logarithmic transformation to the data in this study and remaximized the likelihoods for models 1 through 4. The approximate chi-square statistics were 11.76 ($p < 0.01$, $df = 1$), 0.43, and 11.91 ($p < 0.01$, $df = 2$) for model 2 versus model 1, model 3 versus model 1, and model 4 versus model 1, respectively. The log transformation did remove the significant major-locus genotype effect on the rate of increase of LDL with age. However, the transformation did not remove the highly significant differences in individual-specific environmental effects between those with the mutant allele and those with the normal aa genotype. We question the assumption of homogeneity of phenotypic variance among genotypes that has been made by those conducting quantitative genetic studies. This assumption is certainly not true for all traits influenced by genes at a major locus. Beginning with a model that allows for testing of the assumption of homogeneity of the genotype effect throughout the sample enables one to explore a more realistic set of models.

In conclusion, information about age-specific genotype effects contributes to identification of high-risk individuals, and to statistical support for the major locus, as well as to increasing the understanding of gene action. In this study of a known Mendelian disorder (FH), the predicted effect of the mutant allele is to elevate LDL levels early in life, to contribute to differences in phenotypic variance. The commonly used log transformation did not remove the genotype differences for the first and last

effects. Models, such as the one presented here, that incorporate genotype-specific concomitant effects have the potential to more realistically represent the relationship between genotype and phenotype for many human quantitative traits than do models that assume these effects do not exist.

ACKNOWLEDGMENTS

We thank T. Berry for technical assistance and Dr. M. Boehnke, Dr. T. Burns, R. Kuick, and D. Odenheimer for helpful discussions. We sincerely thank Dr. Sandra Hasstedt for her contributions to the development of Pedigree Analysis Package (PAP) and her assistance in its implementation.

This work was supported in part by grant HL-24481 from NIH and DOE contract E(2828), Health and Welfare Canada (604-7-759 and 6605-1658-53), Medical Research Council of Canada (MT-5427), and the Quebec Heart Foundation.

REFERENCES

- Beaty TH, Self SG, Chase GA, Kwitrovich PO (1983): Assessment of variance components models on pedigrees using cholesterol, low-density and high-density lipoprotein measurements. *Am J Med Genet* 16:117-129.
- Boerwinkle E, Turner ST, Sing CF (1984): The role of the genetics of sodium lithium countertransport in the determination of blood pressure variability in the population at large. In Brewer GJ (ed): "The Red Cell: Sixth Ann Arbor Conference." New York: Alan R Liss.
- Bucher KD, Schrott HG, Clarke WR, Lauer RM (1982): The Muscatine cholesterol family study: Distribution of cholesterol levels within families of probands with high, low and middle cholesterol levels. *J Chron Dis* 35:385-400.
- Burns TL (1982): Sampling considerations for the determination of genetic transmission mechanisms in quantitative traits. PhD Thesis, University of Michigan.
- Cannings C, Thompson EA (1977): Ascertainment in the sequential sampling of pedigrees. *Clin Genet* 12:208-212.
- Cannings C, Thompson EA, Skolnick MH (1978): Probability functions on complex pedigrees. *Adv Appl Prob* 10:26-61.
- Cannings C, Thompson EA, Skolnick MH (1980): Pedigree analysis of complex models. In Mielke JH, Crawford MH (eds): "Current Developments in Anthropological Genetics, Volume 1." New York: Plenum Press, pp 251-298.
- Criqui MH, Frankville DD, Barrett-Connor E, Klauber MR, Holdbrook MJ, Turner JD (1983): Change and correlations of change in high and low density lipoprotein cholesterol after 6 years: a prospective study. *Am J Epidemiol* 118:52-59.
- Darlu P, Lalouel JM, Henrotte JG, Rao DC (1983): A genetic study of red blood cell zinc concentrations in man. *Hum Hered* 33:311-318.
- Davignon J, Lussier-Cacan S, Gattereau A, Moll PP, Sing CF (1980): Studies of a large French-Canadian kindred with two forms of hyperlipoproteinemia. In Gotto AM, Smith LC, Allen B (eds): "Atherosclerosis V." New York: Springer-Verlag, pp 255-259.
- Davignon J, Lussier-Cacan S, Gattereau A, Moll PP, Sing CF (1983): Interaction of two lipid disorders in a large French-Canadian kindred. *Arteriosclerosis* 3:13-22.
- Davis CE (1976): The effect of regression to the mean in epidemiologic and clinical studies. *Am J Epidemiol*. 104:493-498.
- Elston RC, Namboodiri KK, Glueck CJ, Fallat R, Tsang R, Leuba V (1975): Study of the genetic transmission of hypercholesterolemia in a 195 member kindred. *Ann Hum Genet* 39:67-87.
- Elston RC, Namboodiri KK, Nino HV, Pollitzer WS (1974): Studies on blood and urine glucose in Seminole Indians: indications for segregation of a major gene. *Am J Hum Genet* 26:13-24.
- Elston RC, Stewart J (1971): A general model for the genetic analysis of pedigree data. *Hum Hered* 21:523-542.

- Fredrickson DS, Levy RI, Lees RS (1967): Fat transport in lipoproteins—an integrated approach to mechanisms and disorders. *N Engl J Med* 276:32–44, 94–103, 148–156, 215–226, 273–281.
- Gagne C, Moorjani S, Brun D, Toussaint M, Lupien PJ (1979): Heterozygous familial hypercholesterolemia: relationship between plasma lipids, lipoproteins, clinical manifestations and ischaemic heart disease in men and women. *Atherosclerosis* 34:13–24.
- Goldstein JL, Brown MS (1983): Familial hypercholesterolemia. In Stanbury JB, Wyngaarden JB, Goldstein JL, Brown MS (eds): “The Metabolic Basis of Inherited Diseases.” New York: McGraw Hill.
- Harlan WR, Oberman A, Carlson S, Mitchell RE: A longitudinal study of serum lipoproteins. *Am J Epidemiol*, in press.
- Hasstedt SJ, Cartwright PE (1979): PAP. Pedigree Analysis Package. Technical Report No. 13, Department of Medical Biophysics and Computing, University of Utah.
- Hasstedt SJ, Cartwright PE, Skolnick MH, Bishop DT (1979): PAP: A Fortran program for pedigree analysis. *Am J Hum Genet* 31: 135a.
- Hasstedt SJ, Meyers DA, Marsh DG (1983a): Inheritance of immunoglobulin E: genetic model fitting. *Am J Med Genet* 14:61–66.
- Hasstedt SJ, Wilson DE, Edwards CQ, Cannon WN, Carmelli D, Williams RR (1983b): The genetics of quantitative plasma Lp(a): analysis of a large pedigree. *Am J Med Genet* 16:179–188.
- Lalouel JM (1979): GEMINI. A Computer Program for Optimization of General Nonlinear Functions. Technical Report No. 14, Department of Medical Biophysics and Computing, University of Utah.
- Lalouel JM, Rao DC, Morton NE, Elston RC (1983): A unified model for complex segregation analysis. *Am J Hum Genet* 35:816–826.
- Lipid Research Clinics Population Studies Data Book: Volume I (1980): The Prevalence Study. NIH Publication No 80-1527.
- McKusick VA (1983): “Mendelian Inheritance in Man. Catalogs of Autosomal Dominant, Autosomal Recessive and X-linked Phenotypes.” Baltimore: The John Hopkins University Press.
- McKusick VA (1984): The human gene map 15 November 1983. *Clin Genet* 25:89–123.
- McGue M, Laskarzewski P, Rao DC, Glueck CJ (1983): The Cincinnati Lipid Research Clinic family study: commingling in the distributions of lipids and lipoprotein concentrations. *Hum Hered* 33:223–230.
- Moll PP, Berry TD, Weidman WH, Ellefson R, Gordon H, Kottke B (1984): Detection of genetic heterogeneity among pedigrees through complex segregation analysis: an application to hypercholesterolemia. *Am J Hum Genet* 36:197–211.
- Morton NE, Gulbrandsen CL, Rhoads GG, Kagan A, Lew R (1978): Major loci for lipoprotein concentrations. *Am J Hum Genet* 30:583–589.
- Morton NE, MacLean CJ (1974): Analysis of family resemblance. III. Complex segregation of quantitative traits. *Am J Hum Genet* 26:489–503.
- Murphy EA (1979a): Segregation of noisy Mendelian traits and the effect of age-dependence: a prolegomenon. *Am J Med Genet* 4:173–190.
- Murphy EA (1979b): Quantitative genetics: a critique. *Soc Biol* 26:126–141.
- Myant NB (1981): “The Biology of Cholesterol and Related Steroids.” London: William Heinemann Medical Books Ltd.
- Namboodiri KK, Elston RC, Hames C (1977): Segregation and linkage analysis of a large pedigree with hypertriglyceridemia. *Am J Med Genet* 1:157–171.
- Neter J, Wasserman W (1974): “Applied linear statistical models: regression, analysis of variance and experimental designs.” Homewood, Illinois: Richard D Irwin.
- Sabine JR (1977): “Cholesterol.” New York: Marcel Dekker.
- Sokal RR, Rohlf FJ (1969): “Biometry: The Principles and Practice of Statistics in Biological Research.” San Francisco: WH Freeman and Company.
- Webber LS, Smoak CG, Srinivasan SR, Berenson GS (1984): Cardiovascular risk factor variables from birth to 7 years of age—the Bogalusa newborn-infant cohort study. *CVD Epidemiology Newsletter of the American Heart Association* 35:44.
- Williams WR, Lalouel JM (1982): Complex segregation analysis of hyperlipoproteinemia in a Seattle sample. *Hum Hered* 32:24–36.
- Vogel F, Motulsky AG (1979): “Human Genetics: Problems and Approaches.” New York: Springer-Verlag.