ANALYSIS OF GENETIC AND ENVIRONMENTAL SOURCES OF VARIATION IN SERUM CHOLESTEROL IN TECUMSEH, MICHIGAN—VI. A SEARCH FOR GENOTYPE BY ENVIRONMENT INTERACTION

JOHN D. ORR[1], CHARLES F. SING[1] and PATRICIA P. MOLL[1,2]

[1]Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109
and [2]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI 48109, U.S.A.

(*Received 9 December* 1980)

Abstract—We have examined 12 unlinked polymorphic genetic marker systems and found 6 which each identified small but statistically significant differences in normal serum cholesterol levels which did not change significantly with age. The effects associated with 4 of the 6—ABO, haptoglobin, Gm and secretor—were consistent in males and females and combined additively to define effects which were also homogeneous over age. There was a marker phenotype by age interaction effect on cholesterol variance in females but not in males. This study supports the hypothesis that a number of polymorphic genes, each with a small additive effect on cholesterol level, play an important role in determining cholesterol variability.

INTRODUCTION

THE LEVEL of serum cholesterol has been found to be one of the primary indicators of risk to coronary heart disease in Caucasian populations, along with blood pressure, blood glucose and smoking [1-4]. We have conducted a number of studies in the Tecumseh, Michigan, population to estimate the relative contributions of shared genes and shared environments to the observed aggregation of cholesterol levels in families. After adjustment for variability in age, replicate measurements and socioeconomic status, we have estimated that 58.5% (SE = 3.8%) of the variance of ln cholesterol in the population is associated with unspecified genetic factors [5]. We found that 9.8% (SE = 3.1%) of the variance of parental values and 7.7% (SE = 3.3%) of the offspring variance were attributable to variability among household environments. The relative contribution of the shared genetic and shared household environmental factors to familial aggregation in Tecumseh is remarkably similar to that estimated by Rao and his colleagues [6] in a study of Japanese-Hawaiians.

A question of major interest in genetic epidemiology is whether the genetic effects which contribute to phenotype variability combine additively with the effects of environment. The implications of genotype by environment ($G \times E$) interaction in interpreting genetic and environmental determinants of quantitative variability are discussed by a number of investigators [7-11]. To answer the question, one must have measures of both the relevant environments and the genetic effects involved in determining the phenotype. Neither is easy to obtain in a large epidemiological survey. In an effort to shed light on the contribution of $G \times E$ interaction to cholesterol variability in Tecumseh, we selected age as an index of a number of environments which might be expected to change over the lifetime of the individual; e.g. maturation, diet, physical activity, stress and socioeconomic status may all influence the expression of genes which are involved in determining lipid levels. Specific gene effects could be measured by inherited variations in gene products directly involved in cholesterol metabolism or by polymorphic genetic blood

and serum markers which are closely linked to such loci. Unfortunately, the only loci which have been found to be directly involved have mutant alleles which are rare in frequency [12]. However, a number of studies have found associations between cholesterol levels and certain polymorphic genetic markers, most notably ABO and secretor [13, 14]. The consistency of observed effects among studies and between sexes within studies suggests that the loci identified by these genetic markers may either themselves be involved in cholesterol metabolism or be very closely linked to the loci which are.

A marker phenotype by age ($MP \times A$) interaction could be expressed by different relationships among the means or among the variances estimated for the marker phenotypic classes at different ages. Up to now, all studies examining the relationships between cholesterol and genetic markers have been conducted using individuals representing a limited range of ages, with the exception of our earlier study [14], which was based on age-adjusted data. To pursue the question of whether marker phenotype effects are consistent for all ages, we studied changes in both the mean and variance of ln cholesterol over the age range 4–92.

SAMPLE AND ANALYTICAL STRATEGY

The data under study are from the Tecumseh, Michigan, Community Health Study, which has been described previously [15–17]. A description of available serum cholesterol data and the method of laboratory analysis have appeared elsewhere [18]. A total of 3095 males and 3227 females had age and non-fasting serum cholesterol recorded between 1963 and 1966 and were typed for each of 12 unlinked polymorphic genetic blood and serum markers. The systems included Gm, Gc, haptoglobin, MNS, Rh, Duffy, P, ABH secretion, Lewis secretion, Kell, Kidd and ABO. The phenotypes considered and their relationships to single-locus genotypes are described in Table 1 of a previous study [14]. Individuals in this sample ranged in age from 4 to 92. Skewness in the frequency distribution of cholesterol [18] made it appropriate to transform all values by the natural logarithm (ln).

Our previous studies [18] established that a third-order polynomial best described the relationship of ln cholesterol and age. Figure 1 shows the best-fitting third-order regressions of ln cholesterol on age for males and females. Because the regression relationship in males was significantly different from that in females, we considered the sexes separately for the study of marker effects on cholesterol prediction at different ages.

A study of cholesterol variability by Moll et al. [19] suggested that segregation at one or more loci with a major effect and extreme environmental effects associated with certain households are likely to be present in the Tecumseh population. Investigations of the families of hypo- and hypercholesterolemic individuals suggested that they represent heterogeneous contributions of genetic and environmental etiologies which present a different kind of research problem than arises with the normocholesterolemic population.
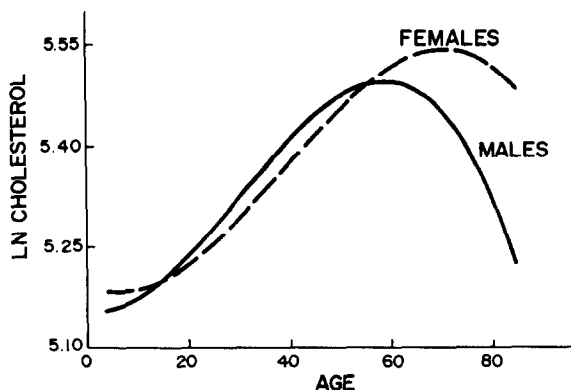

Fig. 1. Third-order regressions of ln cholesterol on age for males and females in Tecumseh, Michigan.

TABLE 1. ESTIMATES OF PARAMETERS OF THE DISTRIBUTIONS OF ADJUSTED LN CHOLESTEROL

Distribution	Males		Females		
	1	2	1	2	3
Percent of individuals included	0.96	99.04	0.63	97.37	2.00
Ln cholesterol mean	4.822	5.302	4.849	5.303	5.655
(antilog—mg%)	(124.2)	(200.7)	(127.6)	(200.9)	(285.7)
Standard deviation*	0.1640			0.1586	

* The variances of the distinct distributions were not found to be different at the 0.10 level of significance in either sex.

Because our aim was to investigate the interaction of genetic and environmental effects on *normal* cholesterol variability, we chose to eliminate from our sample those individuals representing extreme effects. We began by examining the sample defined earlier [14], using the maximum likelihood method of Day [20] to determine if more than one frequency distribution was necessary to describe ln cholesterol for males and/or females in the Tecumseh population. Table 1 gives the estimates of the parameters of the best-fitting distributions for each sex. About 1% of the sample of males belongs to a second distribution with a lower mean cholesterol. Three distinct distributions of females were detected. The distributions with the highest and lowest means included a total of 2.6% of the female sample. To assign an individual to a distribution, we divided the areas of distribution overlap into 6 equal segments. The estimates of the parameters of the distributions given in Table 1 were used to define each segment, assuming a mixture of normal probability functions. We randomly selected from each segment those individuals to be assigned to the distribution of extreme cholesterol values. Individuals identified as being in the distributions representing the extreme values were removed from further consideration, along with every member of their household. Although eliminating such households may not totally remove the major gene and extreme environmental effects from our sample, in the households that remain such effects are not expressed, if they do exist. Approximately 5.4% of the males and 5.7% of the females resided in the households identified as having extreme effects, leaving 2929 males and 3043 females to be analyzed for MP × A interaction. The third-order polynomials for all data, given in Fig. 1, were not significantly altered in this subset of data.

To identify MP × A interaction effects on the level of cholesterol, we investigated the variability of the third-order regression of ln cholesterol on age among marker phenotypes using the analysis of covariance. The traditional error mean squares used for testing the hypotheses of equal regression slopes and equal age-adjusted phenotype means would underestimate the expected error variances because members of the same household share genes and/or environmental effects. An appropriate error term was computed as a function of mean squares [21] obtained by carrying out the nested analysis of variance we described previously [14] on the ln cholesterol values adjusted for a third-order age regression. In order to accommodate this nested design, when more than one phenotype for a given system was present in a household, only one was chosen (at random without regard to the number of individuals in the household with that phenotype) to represent the household in the analysis of marker effects. Each system was considered independently; as a result, the composition of the subsamples differed somewhat among systems. The redefinition of the sample to obtain a valid error term reduced the amount of data available by between 8 and 11% for dominant marker systems and between 14 and 19% for codominant systems.

Before proceeding to an analysis of covariance to detect MP × A interaction and marker phenotype effects on lipid levels for each of the 12 systems in this study, we examined the distribution of the variance of ln cholesterol among phenotype and age classes. The variance is of interest for two reasons. First, two assumptions of the analysis

of covariance are that the variance about the regression curve for each phenotype is constant over all ages and that the variances of ln cholesterol within marker phenotype classes are equal. Failure for these assumptions to be true would invalidate the conclusions one might draw from an analysis of covariance. Second, heterogeneity of variance is another measure of the failure of all genotypes to respond similarly to a population of environments. If the variance is homogeneous over age for each marker phenotype, this type of genotype by age interaction would not be indicated. The same conclusion could be drawn even if the variances were heterogeneous over age, provided the pattern of change in variance was the same in each phenotype. However, if the pattern of change in variance differed significantly among marker phenotypes, we would have evidence for a genotype by age interaction effect on cholesterol variability.

To examine the distribution of the variance among 5-yr age cohorts and marker phenotypes, we made use of a modification of the procedure of Bartlett [22] to carry out a two-way analysis of heterogeneity of variance. We calculated Bartlett's chi-square statistic for heterogeneity of variance among R age classes, among C marker phenotypes, and among the $R \times C$ age-phenotype combinations. A chi-square for testing the $MP \times A$ interaction effect on variance was calculated by subtracting the age and marker phenotype chi-squares (with $R - 1$ and $C - 1$ degrees of freedom, respectively) from the chi-square computed on the age-phenotype cells (with $N - 1$ degrees of freedom, where $N = R \times C$). The degrees of freedom for the test of interaction were taken to be $N - R - C - 1$. This procedure was followed for each marker system in each sex.

In the absence of heterogeneity of variance, the investigation of marker effects on mean cholesterol levels over age began with a test of the hypothesis that the regression curves for each phenotype were parallel. If we failed to reject this hypothesis (meaning there was no indication of $MP \times A$ interaction), we proceeded to test the hypothesis that the age-adjusted phenotype means were equal. If we found that the age-adjusted means were significantly different from one another, we concluded that the marker system under study identified a genetic effect on cholesterol and that this effect was consistent over all ages. If we rejected the first hypothesis, that the regression relationships were the same, we concluded that the marker phenotypes predicted cholesterol differently at different ages, in which case there was evidence for a genotype by age (environment) interaction. A model appropriate for the interpretation of $G \times E$ interaction will be included with the discussion of the statistical analyses presented below.

RESULTS

Heterogeneity of variance

For males, we found that the chi-square measure of the $MP \times A$ interaction effect on variance was significant at the 0.05 level of probability only for the MN system. Ignoring age, Kell was the only system for males which identified statistically significant differences in variance among marker phenotypes. In view of the fact that we analyzed 12 independent marker systems, it is reasonable to dismiss the single significant result for each type of test (of interaction and of marker effects) as a type I error. Neither $MP \times A$ interaction effects on variance nor heterogeneity of variance among marker phenotypes appears to be a major characteristic of the male data from Tecumseh.

In contrast to the findings in males, $MP \times A$ interaction effects on variance were significant at the 0.05 level of probability for the ABO, haptoglobin, secretor, Kidd and P systems for females. In addition, the test for interaction was significant at the 0.08 level for the Gc, Duffy and Rh D systems, suggesting that in as many as 8 of the 12 marker systems we studied, the interaction of genotype with environmental factors indexed by age may play an important role in the determination of cholesterol variability for females.

If there were no $MP \times A$ interaction effect on variance, we would expect variability within each marker phenotype to change approximately the same amount from one 5-yr age cohort to the next. The detection of an interaction effect in females implies that the
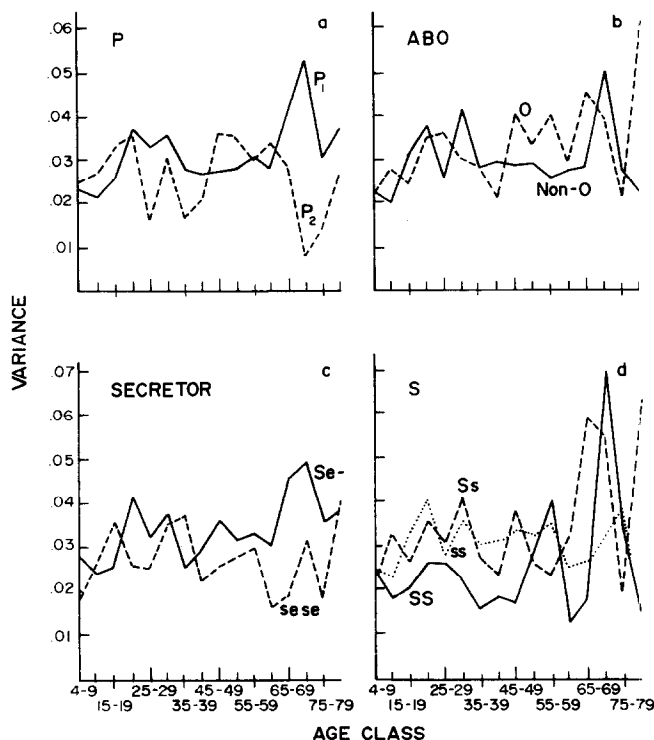
Fig. 2. Variances of ln cholesterol in phenotypes of 4 polymorphic marker systems in 5-yr age cohorts for females.

changes in the variance from one age cohort to the next differ among the phenotypes of a marker system. The plots of the variance against age class within each marker phenotype suggest that interaction effects are, for the most part, attributable to changes in rank of the marker phenotypes at 15, 45 or 65 yr of age for all but the Kidd and Gc systems. To illustrate, the plots for P, ABO, and secretor are given in Figs 2a, 2b and 2c, respectively. The P_2 phenotype predicts lower variance in the 20–44 and 65–92 age ranges, whereas the O and secretor phenotypes predict greater variance in the 45–69 and 40–79 age ranges, respectively. For statistical comparisons, we pooled the 16 5-yr age cohorts into four: 4–14, 15–44, 45–64 and 65–92. We then re-examined the differences among variances estimated for each of the marker phenotypes within the pooled cohorts for all systems. Statistically significant (at the 0.05 level of probability) differences in the variance among individuals aged 15–44 having different marker phenotypes were detected for P, Duffy and S, while the secretor, haptoglobin, Rh D, P and S systems identified significant heterogeneity among phenotypic classes in the 65–92 age category. ABO is the only system where significant phenotypic differences in variance were detected in the 45–64 age class.

There were 4 independent marker systems where no significant interaction was indicated for females. Of these, only for the S system was there statistically significant heterogeneity of variance among phenotypes ignoring age. Figure 2d illustrates the lower variance among females aged 10–45 who are homozygous S.

Since the marginal variances in the 16 age cohorts were based on nearly the same data for the analysis of each system, we investigated the variances of ln cholesterol among age cohorts ignoring marker phenotype. The variance was significantly heterogeneous at the 0.05 level of probability for both sexes. Greater variance was observed between ages 15 and 30 and between 60 and 80. The biological ramifications of the observed heterogeneity of variance will be discussed below.

TABLE 2. RESULTS OF TESTS ($\alpha = 0.05$) FOR MARKER PHENOTYPE BY AGE INTERACTION AND MARKER PHENOTYPE EFFECTS ON CHOLESTEROL

	Interaction with age?		Significant marker phenotype mean effects*		
Marker system	Males	Females	Males	Females	Consistent between sexes?
Gc	No	Yes	No test	No test	No test
Haptoglobin	No	No	—	22 > 1 –	Yes
Gm	No	No	—	a– > bb	Yes
Secretor	No	No	sese > Se-	sese > Se-	Yes
ABO	No	No	non-O > O	—	Yes
Rh E	No	No	E- > ee	—	No
Duffy	No	No	—	$a^- > a^+$	No

* After correction of data for third-order relationship with age.

Heterogeneity of regression on age among marker phenotypes

Because heterogeneity of error variance violates an assumption of the analysis of covariance we have proposed, we proceeded with a weighted analysis which utilized the inverse of the subclass variances [23] to investigate $MP \times A$ interaction effects on mean cholesterol levels. Table 2 summarizes the results of our search for $MP \times A$ interaction and marker phenotype effects. Statistically significant interaction was not found for any of the 12 marker systems in males and only for the Gc system in females. Figure 3 shows that the predicted ln cholesterol of the Gc heterozygote rises nearly uniformly, while the predictions for the two homozygotes change in the same third-order way that the overall population mean changes with age. Out of the 24 independent tests of interaction, it is expected that one would be significant by chance alone; therefore, we chose to interpret the Gc interaction to be a spurious result. Confirmation of this interaction in an independent study, however, would suggest that genetic effects marked by Gc may not be independent of the changes which are occurring with age.

Differences among age-adjusted marker phenotype means significant at the 0.05 level of probability were found for the ABO and Rh E systems in males, for the haptoglobin, Gm and Duffy systems in females, and for the secretor system in both sexes (Table 2). Such differences could not be appropriately tested in females for the Gc system over all ages since $MP \times A$ interaction effects were found to be significant in that case. A significant difference was found among the Gc phenotype means only for those in the age range 4–14. Table 3 gives the adjusted means and standard errors for each phenotype of each of the six marker systems in both sexes. In an effort to establish the nature of the effects which determined the significant heterogeneity of mean cholesterol among marker
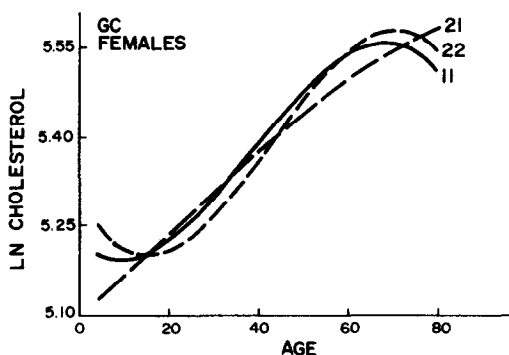

Fig. 3. Third-order regression of ln cholesterol on age for phenotypes of the Gc system for females.

TABLE 3. AGE-ADJUSTED LN CHOLESTEROL MEANS AND THEIR STANDARD ERRORS FOR EACH PHENOTYPE OF THE 6
MARKER SYSTEMS WHERE A STATISTICALLY SIGNIFICANT MARKER PHENOTYPE EFFECT WAS OBSERVED IN AT LEAST ONE
SEX

		Males			Females		
System	Phenotype	Sample size	Adjusted mean Ln Chol ± SE	Antilog	Sample size	Adjusted mean Ln Chol ± SE	Antilog
ABO	A_1	838	5.3458 ± 0.00604	209.7	915	5.3298 ± 0.00564	206.4
	A_2	222	5.3482 ± 0.01271	210.2	223	5.3259 ± 0.01161	205.6
	B	229	5.3441 ± 0.01119	209.4	198	5.3520 ± 0.01232	211.0
	A_1B	52	5.3396 ± 0.02348	208.4	53	5.3060 ± 0.02011	201.5
	A_2B	28	5.2870 ± 0.03364	197.7	28	5.2863 ± 0.03134	197.6
	O	1103	5.3163 ± 0.00542	203.6	1189	5.3263 ± 0.00510	205.7
	non-O	1369	5.3444 ± 0.00480	209.4	1417	5.3306 ± 0.00457	206.6
Haptoglobin	11	378	5.3228 ± 0.00931	205.0	405	5.3220 ± 0.00849	204.8
	21	1201	5.3284 ± 0.00510	206.1	1238	5.3246 ± 0.00483	205.3
	22	837	5.3389 ± 0.00643	208.3	914	5.3447 ± 0.00597	209.5
	1 –	1579	5.3270 ± 0.00451	205.8	1643	5.3240 ± 0.00422	205.2
Gm	aa	233	5.3383 ± 0.01106	208.2	255	5.3339 ± 0.01075	207.2
	ab	1072	5.3338 ± 0.00534	207.2	1119	5.3385 ± 0.00524	208.2
	bb	1173	5.3239 ± 0.00528	205.2	1251	5.3200 ± 0.00487	204.4
	a –	1305	5.3346 ± 0.00485	207.4	1374	5.3376 ± 0.00474	208.0
Secretor	Se –	2022	5.3177 ± 0.00399	203.9	2114	5.3216 ± 0.00385	204.7
	sese	618	5.3477 ± 0.00702	210.1	651	5.3422 ± 0.00647	209.0
Rh E	EE	62	5.3621 ± 0.02393	213.2	64	5.3103 ± 0.01791	202.4
	Ee	679	5.3392 ± 0.00660	208.3	660	5.3181 ± 0.00689	204.0
	ee	1897	5.3204 ± 0.00413	204.5	2056	5.3269 ± 0.00382	205.8
	E –	741	5.3412 ± 0.00645	208.8	724	5.3174 ± 0.00655	203.9
Duffy	Fy a⁺	1826	5.3319 ± 0.00419	206.8	1849	5.3221 ± 0.00408	204.8
	Fy a⁻	782	5.3239 ± 0.00654	205.2	859	5.3359 ± 0.00590	207.7

phenotypes, we used Scheffé's simultaneous contrasts. This analysis identified two classes of phenotypes for each marker system. ABO was divided into O and non-O phenotypes, while in the haptoglobin, Gm and Rh E systems the heterozygote was combined with one of the homozygotes. The means and their standard errors for the combined phenotypes also appear in Table 3. When the six marker systems were each represented by two phenotypic classes, the significant age-adjusted mean differences ranged from 2.9 mg% for Duffy in females to 6.2 mg% for secretor in males.

The estimated changes in mean cholesterol over age for each of the marker phenotypes of the six systems identified are given in Fig. 4. As expected from the covariance analysis, the regressions estimated for the phenotypic classes of a marker system are parallel for every one of the six marker systems for both males and females. That these regressions are parallel illustrates that the gene loci marked by these systems contribute independent of age to the prediction of differences in cholesterol level. In both sexes, the mean age-adjusted ln cholesterol of the nonsecretors was significantly higher than that of the secretors. While the haptoglobin, Gm and ABO age-adjusted contrasts were significant in only one sex, the rank order of the means of the phenotypic classes for those systems was the same in the sex where the differences were not significant. Consistent ranking of the marker effects in both sexes, even if the effects were statistically significant in only one sex, is viewed as evidence that a marked gene effect is being expressed in the Tecumseh population. In the case of Rh E and Duffy, the observed effects were inconsistent between males and females; hence, they could reflect either sex-specific effects or a spurious marker effect in one sex.

Multilocus heterogeneity

Having considered each marker system apart from the others, we proceeded to examine cholesterol levels when the marker systems were considered simultaneously. To do this, we chose the four systems where the age-adjusted means were significantly different
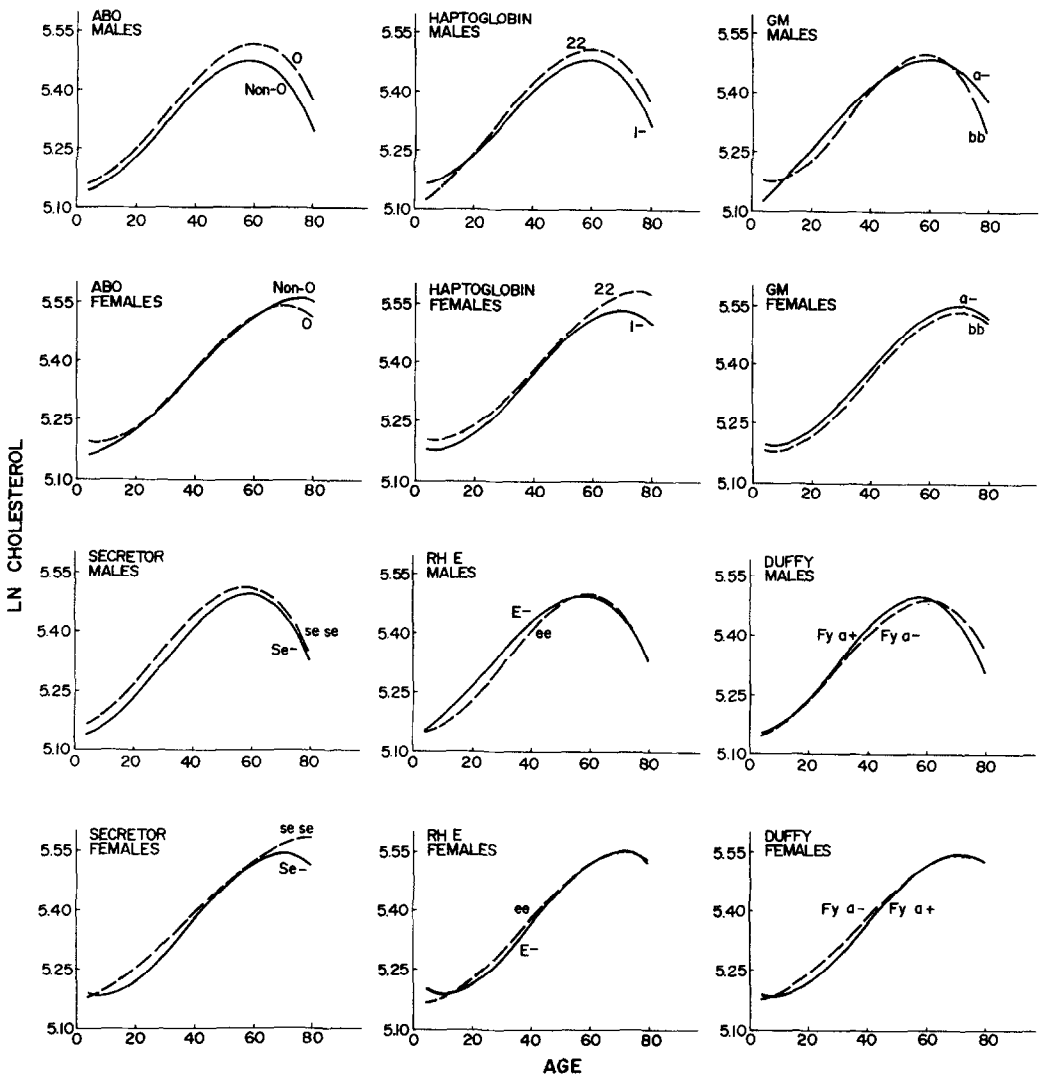
Fig. 4. Third-order regressions of ln cholesterol on age for the phenotypes of 6 systems where a statistically significant marker phenotype effect was observed in one or both sexes.

among phenotypes in at least one sex and ordered the same in both sexes, i.e. ABO, haptoglobin, Gm and secretor. We identified for each system a 'positive marker effect', that is, the phenotypic classes that we associated with higher cholesterol values. Table 3 shows that the positive effects were associated with the non-O, Hp 22, Gm a- and nonsecretor phenotypic classes. An individual's multilocus marker phenotype was defined as the number of positive marker effects that individual had. Figure 5 shows the regressions of ln cholesterol on age for each of these multilocus phenotypes, separately for males and females.

Before considering the statistical properties of the relationship of age with multilocus phenotype means, we examined the distribution of variance among multilocus phenotypes and age classes. Using the two-way analysis of heterogeneity of variance described above, we found a multilocus phenotype by age interaction effect on variance which was significant at the 0.05 level of probability for females. No multilocus marker phenotype had significantly greater or smaller variance than the others over any large age range. For males there was no significant heterogeneity of variance with respect to the number of positive marker effects. Because of the observed heterogeneity of variance, a weighted analysis of covariance (like that used in the analysis of single-locus effects) was used to
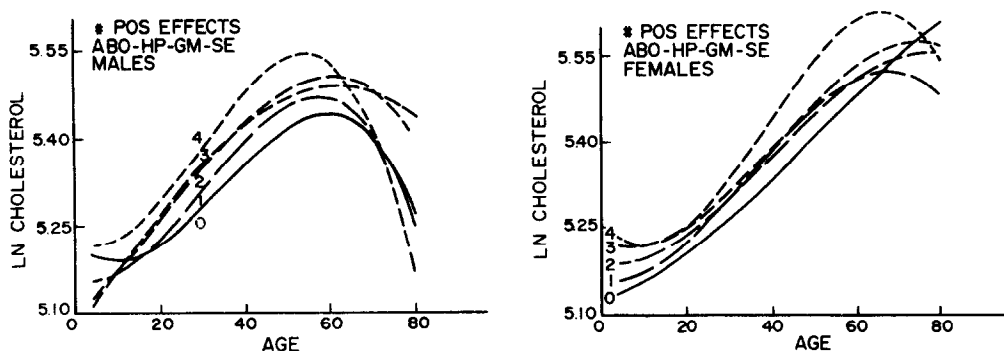
Fig. 5. Third-order regressions of ln cholesterol on age for the multilocus marker phenotypes.

test for multilocus phenotype by age interaction effects on cholesterol levels. This analysis indicated that the regression relationships were not significantly different among multilocus phenotypes at the 0.10 level of probability for either sex; hence, we conclude that there is no evidence for a multilocus phenotype by age interaction effect on means.

The age-adjusted means, which appear in Table 4 along with their standard errors, increased with the number of positive effects in both sexes. A difference of approximately 18 mg% was predicted by these four marker systems in both sexes. The regressions in Fig. 5 show that the rank order of the means of the multilocus phenotypes prevailed from age 15 to 65, with the exceptions that males with 2 positive effects had slightly greater mean cholesterol after age 35 than those with 3 positive effects and females with 1 positive effect had slightly greater mean cholesterol between ages 30 and 55 than those with 2 positive effects. The age-adjusted means of the involved phenotypes were so close together that it is not surprising that their order be reversed at some ages; and, the reversal of order was not great enough in these data to cause a statistically significant multilocus phenotype by age interaction. The increase in mean cholesterol, which was consistent over age, with the number of positive marker effects indicates that these polymorphic genetic markers, or genetic loci which are closely linked to them, are each responsible for a small, additive effect on cholesterol level which is statistically independent of the age of an individual.

DISCUSSION

Our statistical analyses of the distribution of serum cholesterol across age cohorts suggest two conclusions. First, the statistically significant genetic effects on the level of

TABLE 4. AGE-ADJUSTED LN CHOLESTEROL MEANS AND THEIR STANDARD ERRORS FOR MULTILOCUS PHENOTYPES BASED ON THE NUMBER OF POSITIVE MARKER EFFECTS

	No. positive effects	Sample size	Adjusted mean Ln Chol ± SE	Antilog
Males	0	167	5.3140 ± 0.01482	203.2
	1	595	5.3339 ± 0.00711	207.2
	2	720	5.3654 ± 0.00686	213.9
	3	274	5.3672 ± 0.01008	214.3
	4	37	5.4015 ± 0.03038	221.7
Females	0	193	5.3241 ± 0.01276	205.2
	1	645	5.3509 ± 0.00676	210.8
	2	747	5.3554 ± 0.00669	211.7
	3	315	5.3717 ± 0.00934	215.2
	4	44	5.4084 ± 0.01976	223.3

age-adjusted cholesterol identified by four polymorphic blood and serum marker systems that we reported earlier [14] are consistent in their size and direction over all age classes for both sexes. Second, 8 of the 12 marker systems studied identified significant genetic effects on phenotypic variability at specific periods of life in females but not in males. A meaningful discussion of the biological interpretations of these results cannot be undertaken without establishing a concise definition of the population of genetic and environmental effects that are represented in the sample which was subjected to the statistical analysis. Because we do not have direct measures of the genes and the environments which contribute to each individual's cholesterol phenotype, we must construct a mathematical definition of these factors that is based on an appropriate stratification of the sample of phenotypes. Such a definition involves the assignment of a phenotypic value to each genotypic and environmental stratification considered by our study and knowledge about the joint frequency distribution of those values.

We will follow the arguments developed by Comstock and Moll [24] to model the observed effects of age and marker phenotype on the level and variability of serum cholesterol in terms of the contribution of genetic and environmental sources of phenotypic variability. A convenient linear model which describes the contribution of genetic and environmental effects which are identified by the analytical design of this study may be written

$$P_{ijk} = \mu + G_i + E_j + GE_{ij} + e_{ijk}.$$

Here G, E, GE and e are unmeasured random variables that define, respectively, the separate contributions of genotype, the macroenvironment defined by age class, an interaction between the genotype and the age-specific environmental effects, and a random environmental effect within the age-genotype class. The value of each of these variables is taken to be the deviation of the mean of an appropriate subset of individuals from the mean of all individuals of the population under consideration. For example, the expected genetic contribution to the phenotype of the individual with the ith genotype is taken to be the deviation from the overall mean (μ) of the average phenotype of all individuals with the ith genotype weighted by the relative frequency of the ith genotype class. In our study, classes of genotypes are identified by the marker phenotypes. Defined in this way, the value of each genotype depends on the phenotypes it produces and these, of course, depend on the environments experienced. Consequently, the interpretation of the ith genotype value as $P_{i..} = \mu + G_i$ depends on the distribution of the population of genotypic effects with respect to the factors defined by the variables E and e. Before considering the consequences of this definition, we similarly define E_j as the deviation of the jth age class mean from μ and e_{ijk} as the deviation of the value of the kth individual from the mean of the ith genotype in the jth age class studied. Age indexes a class of environmental factors whose effects are distributed among age cohorts. In this study the jth environmental value, $P_{.j.} = \mu + E_j$, is taken to be the average of those individuals who experience the jth combination of macroenvironmental outcomes that vary among the 5-yr age classes. Diet, exposure to pollutants, stress, physical activity and any other factors which vary as a function of age may be candidates.

The G_i and E_j effects are independent measures of genotype and environment, respectively, only if their frequency distributions are uncorrelated; that is, $f(GE_{ij}) = f(G_i) \times f(E_j)$. If certain genotypes are found in certain age groups more (or less) often than expected, it is not possible to contrast genotype values in different age classes in a meaningful way. Simply, genotype values may compare differently in different age strata, not because the effects of the genotypes are different, but because they are differentially represented in the strata being compared. The assumption of independence of the frequencies of the G_i and E_j effects in this study is supported by an earlier analysis [25] which found that estimates of both the phenotype and the allele frequencies for each of the 12 markers considered here were not significantly heterogeneous among age classes in these data. With this assurance it is reasonable, then, to proceed to define an interaction effect GE_{ij} as the failure of the genetic effect estimated by the ith genotype

value and the environmental effect estimated by the average of all individuals in the jth age class to explain the average phenotype of those individuals with the ith genotype who fall in the jth age class. The distribution of $G \times E$ interaction effects explains the variability among the means of subclasses defined by the ith genotype and the jth macroenvironment $(P_{ij.})$ not accounted for by the distribution of the genotypic and environmental values about the overall mean. An exact definition of the variabile e_{ijk} remains. For the analysis presented above, e_{ijk} represents the deviation of the kth individual from $P_{ij.} = \mu + G_i + E_j + GE_{ij}$. With this model the total phenotypic variance of ln cholesterol can be partitioned thusly:

$$\sigma_p^2 = \sigma_G^2 + \sigma_E^2 + \sigma_{GE}^2 + \sigma_e^2$$

where the subscripts define the associated variances. The phenotypic variance for the ith genotype and the jth environment is defined

$$\sigma_{P_{ij}}^2 = \sigma_{e_{ij}}^2$$

and is attributable to those genetic and microenvironmental differences not accounted for by the definition of G_i and E_j used here.

Environments can interact with genotypes to cause several different types of effects on the resultant phenotype. One type of $G \times E$ interaction effect, and that which is most often examined, is that on location, or the average value of the phenotype. In terms of the model, when interaction is present, the interaction effect GE_{ij} is not identically zero for all combinations of genotypes and all environments; equivalently, σ_{GE}^2 is not equal to zero. Heterogeneity of $\sigma_{e_{ij}}^2$ among subclasses defined by the ith genotype and the jth environment is a second type of interaction between genotype and environment. Murphy [26] points out the importance of utilizing the second and higher moments of data and laments that such analyses are too infrequently called upon to understand more fully the biological basis of a trait.

Variability among individuals within each marker phenotype-age subclass, $\sigma_{e_{ij}}^2$, may be attributable to the expression of unmeasured genetic and microenvironmental factors which are randomly distributed within each marker-age subclass. The heterogeneity of this variance among subclasses may be due to the involvement of genetic and/or environmental factors that are randomly distributed in all subclasses but expressed only in those individuals with specific marker phenotypes and belonging to specific age classes. The differential response of marked genotypes to the microenvironments in different age classes is manifested by statistically significant heterogeneity of variance among marker phenotype-age classes. This type of effect of genotype on phenotypic variability has been defined by Lerner [27] as differential homeostasis (where the stability of phenotype variability in different macroenvironments depends on the genotype) and by Bradshaw [28] as differential phenotypic plasticity. Heterogeneity of variance among subclasses might also be attributable in part to a correlation between unmeasured genetic and random microenvironmental effects and the marker phenotype-age strata. Such a correlation is impossible to separate from $G \times E$ interaction in a cross-sectional survey of this type. That there is heterogeneity of variance among age strata suggests that unmeasured genetic and microenvironmental factors may not be randomly distributed in these data.

The formal definitions of the effects specified in the model above should leave no doubt that when one speaks of the effect of a particular genotype or of a particular environment, it is in the context of a particular population of environments and a particular population of genotypes, respectively. Furthermore, the genetic interpretation of the difference between two genotype values has little meaning when there is $G \times E$ interaction unless one specifies the environmental stratification. That is, when there is nonadditivity of genotype and environment, the effect of a genotype is defined as a function of both its contribution to the average phenotype over all environments and its differential response as it interacts with specific subsets of environments. In addition, it must be emphasized that such an interpretation holds only when there is no correlation between genotype and environmental values.
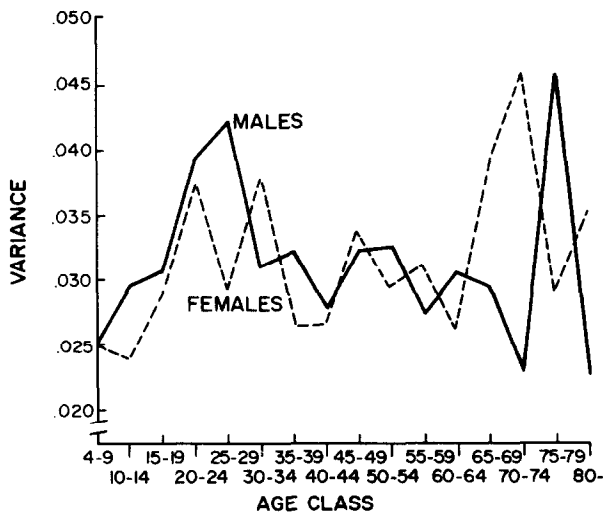
Fig. 6. Variances of ln cholesterol in 5-yr age cohorts for males and females, ignoring markers.

Figure 6 shows the variances for each 5-yr age cohort by sex, ignoring marker pheno-type. In the Tecumseh population we have found statistically significant heterogeneity of variance among age classes (ignoring marker phenotypes) in both sexes. Heterogeneity of variance of cholesterol among age classes, while rarely discussed, has been widely reported [29–32]. However, comparing the patterns of variance over age from the above studies to those we observed is inappropriate since the cholesterol data from the other studies were not ln-transformed. In these other data sets the variance of cholesterol values is highly correlated with the mean, a fact we reported earlier for the Tecumseh data [18]. One noteworthy observation from the aforementioned studies is that in 3 of the 4 [30–32] there is an increase in variance of cholesterol which is disproportionately large compared to the increase in mean cholesterol for females around age 20. This is consistent with the finding in our study that the variance of ln cholesterol for females is high in the age cohort 20–24. Heterogeneity of variance among age classes might be attributable to unmeasured genes whose effects are expressed only at certain develop-mental stages in life or environmental conditions which are experienced only by individuals in certain age cohorts.

Up to age 45, the variance for males exceeded that for females, except in the age class 30–34. (In two of the age cohorts, 10–14 and 25–29, the differences were statistically significant.) After age 45, there was no pattern of relationship between the variances in the two sexes, although in the age class 70–74, females had significantly greater variance than males. Similar results have been observed in a population of working Canadians [31], in a combination of 10 populations from the LRC studies [29], and in families from Muscatine, Iowa [32]. In these studies, however, the variance for females exceeded that for males in the age class 20–24, rather than age 30–34, as we observed in Tecumseh.

When we examined the variances over age within marker phenotypes, we found additional differences between the sexes. The pattern of change in variance over age differed significantly among phenotypes in only one of the 12 independent marker systems we considered for males; for females 8 of the 12 systems showed evidence of $MP \times A$ interaction. What environments varying with age might be interacting with genotype to affect the stability of serum cholesterol? Since the significant $MP \times A$ interaction effects we observed were found only in females, the environments which are combining with different genotypes to cause heterogeneous variation are likely to be unique to females. One such environmental factor may well be sex hormones. Many investigators have found significant associations between mean cholesterol levels and hormone usage, although the results have not been consistent and most studies have been based on small

samples. Furman [33] and Bierman [34] review the work in this area. In the LRC studies of 10 diverse North American populations, women up to age 49 using oral contraceptives and estrogen preparations had greater plasma cholesterol levels than those not taking sex hormones [29] and the increase in cholesterol level associated with hormone usage was consistent over the 10 populations [35]. While we are aware of no studies which examine the variance of hormone levels over age or cholesterol variance at different levels of exogenous hormones and while we have no information on hormone levels in Tecumseh, it seems reasonable to hypothesize that individual variability of sex hormone levels might be involved in altering the stability of cholesterol levels in specific genotype-age classes. At the very least, such a hypothesis merits further investigation when the appropriate data are available.

The results obtained from our study of means show that there are significant effects on mean serum cholesterol which are associated with marker phenotypes and that they are independent of macroenvironments which vary with age, i.e. the differences in cholesterol levels among marker phenotypes remain constant throughout the age range 4–92. The marker phenotype effects we observed are consistent with those effects seen in our earlier study [14] and other studies where the effects were not examined over age. Many studies have found the A-carrying phenotypes of the ABO system to be associated with higher levels of cholesterol than non-A phenotypes, while a number of reports indicate that nonsecretors have greater cholesterol levels on average than secretors. Mayo et al. [36] suggest that the association between ABO phenotypes and serum cholesterol 'is possibly related to the role of cholesterol in the erythrocyte membrane, of which the ABO antigens form a part'.

Upon examination of the multilocus phenotypes defined above, we found that mean cholesterol increased with the number of positive marker effects. In addition, we found that, as in the single-locus markers, there was no significant multilocus phenotype by age interaction of means. Although the single-locus marker phenotype differences that we have identified were small, the fact that these differences were consistent over age, the fact that they were consistent with differences observed in other populations, and the fact that together they additively predicted larger differences combine to provide strong evidence that genetic variation is indeed responsible for at least a fraction of the normal cholesterol variability in human populations.

In addition to genes with small effects, such as those we examined above, major genes, such as the gene governing familial hypercholesterolemia (FH), may contribute to cholesterol variability. Figure 7 presents the predicted regressions of ln cholesterol on age for hypercholesterolemic and normal males and females who are members of a 195-member kindred described by Elston and his colleagues [37]. Figure 7 shows that the mean cholesterol of males with the hypothesized FH gene was much greater than that of normal males, but the changes with age in the two groups did not differ greatly. Females with the hypothesized FH gene had mean cholesterol which was much higher than that of normal females at very young ages, and their mean cholesterol increased significantly more rapidly with age than normal females.

Figures 5 and 7 provide a comparison of two types of genetic effect which are important in determining cholesterol variability. The rare FH gene causes serum cholesterol to reach extremely high levels, and there is some preliminary indication (from the females) that a genotype by age interaction may be present. The genes identified by the unlinked polymorphic genetic markers we have studied are associated with small, significant differences in cholesterol levels which are consistent over all ages and combine to define larger differences which are also consistent over all ages. We interpret this consistency over age to mean that any of the macroenvironments which are changing with age are not directly involved in modifying the differences in the small gene effects.

It appears, from consideration of the complexity of the biochemistry of lipoprotein metabolism and the consistency of results from a number of genetic epidemiological studies [38], that we might expect a large number of genetic loci to be involved in modulating serum cholesterol levels among randomly chosen individuals in human
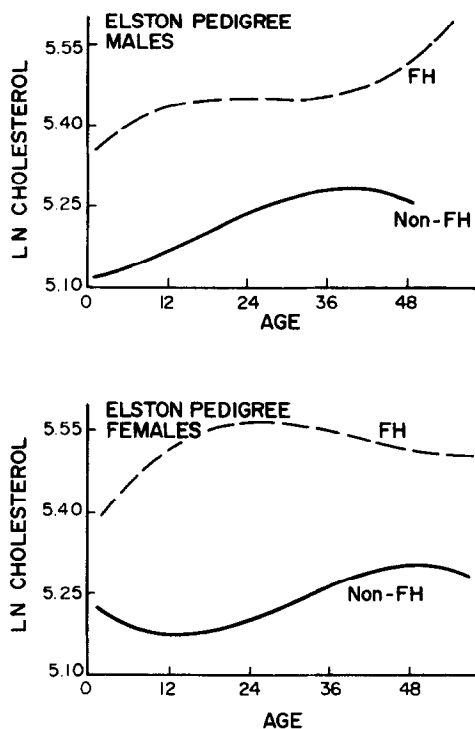
Fig. 7. Third-order regressions of ln cholesterol on age for individuals with familial hypercholes-
terolemia (FH) and those who do not have FH from the 195-member kindred described by Elston
et al. [37].

populations. Of the 12 polymorphic genetic markers we studied, that as many as 6
predicted small differences in serum cholesterol levels (while we know of only a few loci
with major effects on cholesterol levels [39]) strongly suggests that a large fraction of the
loci involved in determining serum cholesterol levels determine small effects. It seems
likely, to us as well as to Edwards [40], that there is a distribution of the size of gene
effects, wherein the loci with large effects (the major genes) are few in number, with alleles
rare in frequency, and loci with small and additive effects (the polygenes) are the most
common, and polymorphic.

REFERENCES

1. Epstein FH, Kjelsberg MO: Coronary heart disease in relation to blood pressure and cholesterol levels in
    population studies. In: **Genetics and the Epidemiology of Chronic Diseases.** Neel JV, Shaw MW, Schull WJ
    (Eds) USPHS Bulletin no. 1163. Washington: Govt Print Office, 1965, pp. 265–277
2. Truett J, Cornfield J, Kannel WB: A multivariate analysis of risk of coronary heart disease in Framing-
    ham. **J Chron Dis** 20: 511–524, 1967
3. Dawber TR, Moore FE, Mann GV: Coronary heart disease in the Framingham study. **Am J Pub Health**
    47 suppl: 4–24, 1967
4. Stamler J: Diet-related risk factors for human atherosclerosis: hyperlipidemia, hypertension, hyperglyce-
    mia—current status. **Adv Exp Med Biol** 60: 125–155, 1974.
5. Sing CF, Orr JD: Analysis of genetic and environmental sources of variation in serum cholesterol in
    Tecumseh, Michigan. IV. Separation of polygene from common environment effects. **Am J Hum Genet** 30:
    491–504, 1978
6. Rao DC, Morton NE, Gulbrandsen CL, Rhoads GG, Kagan A, Yee S: Cultural and biological determin-
    ants of lipoprotein concentrations. **Ann Hum Genet** 42: 467–477, 1979
7. Plomin R, DeFries JC, Loehlin JC: Genotype-environment interaction and correlation in the analysis of
    human behavior. **Psychol Bull** 84: 309–322, 1977
8. Eaves LJ: The multivariate analysis of certain genotype-environment interactions. **Behav Genet** 2:
    241–244, 1972

9. Lewontin RC: The analysis of variance and the analysis of causes. **Am J Hum Genet** 26: 400–411, 1974
10. Feldman MW, Lewontin RC: The heritability hang-up. **Science** 190: 1163–1168, 1975
11. Kidd KK, Matthysse S: Research designs for the study of gene-environment interactions in psychiatric disorders. **Arch Gen Psychiatry** 35: 925–932, 1978
12. Goldstein JL, Brown MS: The LDL receptor locus and the genetics of familial hypercholesterolemia. **Ann Rev Genet** 13: 259–289, 1979
13. Morton NE: Genetic markers in atherosclerosis: a review. **J Med Genet** 13: 81–95, 1976
14. Sing CF, Orr JD: Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. III. Identification of genetic effects using 12 polymorphic blood marker systems. **Am J Hum Genet** 28: 453–464, 1976
15. Napier JA: Field methods and response rates in the Tecumseh Community Health Study. **Am J Pub Health** 52: 208–216, 1962
16. Epstein FH, Ostrander LD, Johnson BC, Payne MW, Hayner NS, Keller JB, Francis T Jr: Epidemiological studies of cardiovascular disease in a total community, Tecumseh, Michigan. **Milbank Mem Fund Q** 43: 333–342, 1965
17. Napier JA, Johnson BC, Epstein FH: The Tecumseh Community Health Study. In: **Casebook of Community Study.** Kessler II, Levin ML (Eds). Baltimore: Johns Hopkins Press, 1971, pp. 25–46
18. Sing CF, Chamberlain MA, Block WD, Feiler S: 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, 1975
19. Moll PP, Powsner R, Sing CF: Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. V. Variance components estimated from pedigrees. **Ann Hum Genet** 42: 343–354, 1979
20. Day NE: Estimating the components of a mixture of normal distributions. **Biometrika** 56: 463–474, 1969
21. Sokal RR, Rohlf FJ: **Biometry.** San Francisco: Freeman, 1969, pp. 277–279.
22. Bartlett MS: Properties of sufficiency and statistical tests. **Proc R Soc Lond** A 160: 268–282, 1937
23. Scheffé H: **The Analysis of Variance.** New York: John Wiley and Sons, 1959
24. Comstock RE, Moll RH: Genotype-environment interactions. In: **Statistical Genetics and Plant Breeding.** Hanson WD, Robinson HF (Eds). Washington: National Academy of Sciences—National Research Council Publication 982, 1963, pp. 164–196.
25. Shreffler DC, Sing CF, Neel JV, Gershowitz H, Napier JA: Studies on genetic selection in a completely ascertained Caucasian population. I. Frequencies, age and sex effects, and phenotype associations for 12 blood group systems. **Am J Hum Genet** 23: 150–163, 1971
26. Murphy EA: Quantitative genetics: a critique. **Soc Biol** 26: 126–141, 1979
27. Lerner IM: **Genetic Homeostasis.** New York: John Wiley and Sons, 1954
28. Bradshaw AD: Evolutionary significance of phenotypic plasticity in plants. **Adv Genet** 13: 115–155, 1965
29. Lipid Research Clinics Program Epidemiology Committee: Plasma lipid distributions in selected North American populations: the Lipid Research Clinics Program Prevalence Study. **Circulation** 60: 427–439, 1979
30. Kwiterovich PO Jr, Chase GA, Bachorik PS: The Columbia population study. I. Plasma cholesterol and triglyceride levels. **Johns Hopkins Med J** 143: 32–42, 1978
31. Hewitt D, Jones GJL, Godin GJ, McComb K, Breckenridge WC, Little JA, Steiner G, Mishkel MA' Baillie JH, Martin RH, Gibson ES, Prendergast WF, Parliament WJ: Normative standards of plasma cholesterol and triglyceride concentrations in Canadians of working age. **Can Med Assoc J** 117: 1020–1024, 1977
32. Schrott HG, Bucher KD, Clarke WR, Lauer RM: The cholesterol family study: the Muscatine study. I. Lipid standardization. in preparation
33. Furman RH: Gonadal steroid effects on serum lipids. In: **Metabolic Effects of Gonadal Hormones and Contraceptive Steroids.** Salhanick HA, Kipnis DM, Vande Wiele RL (Eds). New York: Plenum Press, 1969, pp. 247–264
34. Bierman EL: Oral contraceptives, lipoproteins, and lipid transport. In: **Metabolic Effects of Gonadal Hormones and Contraceptive Steroids.** Salhanick HA, Kipnis DM, Vande Wiele RL (Eds). New York: Plenum Press, 1969, pp. 207–218
35. Wallace RB, Hoover J, Sandler D, Tyroler HA, Rifkind BM: Altered plasma-lipids associated with oral contraceptive or oestrogen consumption. **Lancet** 2: 11–14, 1977
36. Mayo O, Weisenfeld SL, Stamatoyannopoulos G, Fraser GR: Genetical influences on serum cholesterol level. **Lancet** 2: 554–555, 1971
37. Elston RC, Namboodiri KK, Glueck CJ, Fallat R, Tsang R, Leuba V: Study of the genetic transmission of hypercholesterolemia and hypertriglyceridemia in a 195 member kindred. **Ann Hum Genet** 39: 67–87, 1975
38. Sing CF, Orr JD, Moll PP: The contribution of genes controlling total serum cholesterol to prediction of cardiovascular disease risk in the general population. In: **Genetic Analysis of Common Diseases: Applications to Predictive Factors in Coronary Disease.** Sing CF, Skolnick M (Eds). New York: Alan R. Liss, 1979, pp. 597–618
39. Fredrickson DS, Goldstein JL, Brown MS: The familial hyperlipoporoteinemias. In: **The Metabolic Basis of Inherited Disease.** Stanbury JB, Wyngaarden JB, Fredrickson DS (Eds). New York: McGraw-Hill, 1978, pp. 604–655
40. Edwards JH: The genetic basis of common disease. **Am J Med** 34: 627–638, 1963