Pages 271–276, Genetic Analysis of Common Diseases:
Applications to Predictive Factors in Coronary Disease

ON THE INDEPENDENCE STRUCTURE THAT EXISTS WITHIN A PEDIGREE WITH AN APPLICATION TO TESTING FOR A MAJOR GENE

J.A. Williamson,[1] D.T. Bishop,[2] S. Tavaré,[2] and C.J. Tavaré[2]

[1]University of Colorado, Boulder

[2]University of Utah, Salt Lake City

INTRODUCTION

Carmelli, Karlin, and Williams (1978) have presented a wide variety of indices for exploring genetic models using data from independent nuclear families. We present a method of analysis designed for data collected in the form of a pedigree. Our goal is to define a statistic for which a known probability distribution is associated, enabling the investigator to say just how significant any results are. However, when the data come from a single or small number of pedigrees, it would appear more difficult to define a statistic which incorporates the inherent structure of the pedigree. In the following paragraphs we make a modest beginning in the direction of defining, on a pedigree, statistics with known distributions under some null hypothesis. In particular, we define a statistic which seems to have some power to detect major genes. Its empirical power is not as great as we would like, but its null distribution is known so that significance levels can be obtained.

THE MODELS

Our models follow Morton and MacLean (1974). The null model is a polygenic pedigree model. The phenotype of the i^{th} sib in the j^{th} sibship is given by

$$X_{ij} = \mu_j + N_{ij} + C_j + R_{ij}$$

where μ_j is a function of the parental breeding values, and N_{ij}, the individual deviation, is distributed as $N(0,\sigma_1^2)$. C_j is the common sib environment in the j^{th} sibship, and R_{ij} is the phenotypic variation due to random environmental conditions. We assume that N_{ij} and R_{ij} together form a sequence of mutually independent random variables, $i=1,2,\ldots n_j$, $j=1,2,3,\ldots$ where n_j is the number of sibs in the j^{th} sibship. The μ_j and C_j, while random, will not be independent from sibship to sibship. We assume that the N_{ij} and R_{ij} are independent of all μ_k and C_k. There is no major-gene component in our null model.

The one-locus two-allele (A,a) major-gene model is described in Table 1.

Table 1.

		Genotype aa	Aa	AA
Phenotype	Mean	μ	$\mu+\theta d$	$\mu+d$
	Variance	γ^2	γ^2	γ^2

THE INDEPENDENCE RESULT

Theorem. For the polygenic model, let X_{mk} be the phenotype of the m^{th} sib in the k^{th} sibship.

Let $\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}$ denote the average phenotype in the j^{th} sibship. Then, for $k \neq j$, X_{mk} and $X_{ij} - \bar{X}_j$ are independent.

Proof: The proof follows from the observation that the terms μ_j and C_j are not present in the expression $X_{ij} - \bar{X}_j$.

We remark that, in particular, the phenotype of the mother (or father) of the j^{th} sibship is independent of each of the $X_{ij} - \bar{X}_j$. Also, if the k^{th} sibship shares one parent with the j^{th} sibship, each of the X_{mk} are independent of each of the $X_{ij} - \bar{X}_j$.

The theorem is useful in that it gives independence results from which test statistics for the polygenic model may be constructed. For example, it follows that the random

variables $S_j^2 = \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2/(n_j - 1)$, j=1,2,3,.... are mutually independent. This fact forms the basis for the statistical discussion that follows.

A STATISTICAL DISCUSSION

If there is a major-gene component present, then two parents, both heterozygous for the gene, would expect to see more phenotypic variability in their offspring than would two parents who are both homozygous for the gene. This idea and the translation of this idea into tests for the presence of a major gene using nuclear family data have been developed independently by Mérat (1968), Fain and Ott (1976), Fain (1978), and Ott (manuscript).

The phenotypic variance of the polygenic model is constant, and so, referring to Table 2, we see that we should be looking for a way to test for lack of homogeneity of variances among the sibships of our pedigree. This is an extension to pedigrees of the method of Fain and Ott.

Table 2.

Parental Matings	Typical Offspring Phenotype under Major-Gene Model	
	Mean	Variance
Aa x Aa	$\mu + d(1+2\theta)/4$	$\gamma^2 + d^2(3-4\theta(1-\theta))/16$
Aa x aa	$\mu + d\theta/2$	$\gamma^2 + d^2\theta^2/4$
Aa x AA	$\mu + d(\theta+1)/2$	$\gamma^2 + d^2(1-\theta)^2/4$
aa x aa	μ	γ^2
aa x AA	$\mu + \theta d$	γ^2
AA x AA	$\mu + d$	$\gamma 2$

If we assume that all the R_{ij} have the same normal distribution, $N(0,\sigma_2^2)$, then tests based on standard normal theory are available. If we let $\sigma^2 = \sigma_1^2 + \sigma_2^2$, then for fixed n_j, $(n_j - 1)S_j^2/\sigma^2$ has a χ^2 distribution with $n_j - 1$ degrees of freedom.

HYPOTHESIS TESTING

Our null hypothesis is the null model as presented above. In addition we assume that the R_{ij} all have the same normal distribution, $N(0,\sigma_2^2)$. From our pedigree we will use only sibships of a constant size, that is, n_j will be the same for all j. If we write $\max(S_j^2)$ and $\min(S_j^2)$ for the largest and smallest, respectively, of the normalized sibship sum of squares, then we define our test statistic to be $H = \max(S_j^2)/\min(S_j^2)$. The distribution for H is found in Table 31, p 202 of Biometrika Tables (1966). Large values of H will cause us to reject the null hypothesis. We would like to say -- in favor of the alternative hypothesis -- that there is a major-gene component present. Technically, we cannot go this far, though we would argue that if it is not skewness in the data that is causing H to have a large value, then it must be the presence of a major-gene component. Computer simulations, using the simulation package developed by Bishop and Cannings (1978), show that the approach just outlined has a very low power to detect the presence of a major gene. This is not, however, surprising. Our independence theorem tells us that looking at S_j^2 for all sibships of a given size in a pedigree is no different than looking at S_j^2 in the same number of independent nuclear family sibships. This nuclear family approach has been found to lack power by Fain and Ott. We must use the pedigree structure to increase the power of our test. Many different sampling schemes are available, one of which is the following.

For each marriage in the pedigree compute the absolute value of the difference between the male and female phenotypes. Order these absolute differences, and place the marriages that correspond to the largest one-third of the absolute differences into the subpedigree. Also include in the subpedigree all the sibships resulting from these marriages. Any marriages between members of two of these first-to-be-included sibships, and the sibships resulting from these marriages, are also to be included in the subpedigree. Choosing parents that differ widely in their phenotypes should yield several first generation sibships that are nearly all heterozygous for the major

gene, if, in fact, it exists. The next generation of sibships should produce the large phenotypic variability. The statistic, H, applied to this subpedigree could, then, detect the presence of a major gene. At this point we emphasize again that our independence theorem allows us to extract our subpedigree in this way without our losing the distribution of H under our null hypothesis.

Unfortunately, obtaining informative sibships from actual pedigrees is not so easy. We cannot always expect to find an adequate number of sibships that are offspring of matings among our first generation subpedigree sibships. With an actual pedigree, we can enlarge the size of the subpedigree by looking at first generation subpedigree sibships that are removed from each other by one actual generation in the complete pedigree. Then include in the subpedigree, marriages between members of the lower sibships and children of the subpedigree sibships that are one generation higher in the actual complete pedigree, and children of these marriages. Then apply H to this subpedigree. The enlarging of the subpedigree can be extended if more sibships are desired. The important thing to keep in mind is that the inclusion of additional sibships cannot be based on S_j^2 values. Our null distribution on H remains valid only if we extend our subpedigree through decisions made based on properties possessed by parents.

The basic idea is, then, as follows. If a major gene is present, we want, without actually looking at S_j^2 values, to extract a subpedigree that will display the sibship variability due to the major gene, and we want to do this with a subpedigree that contains relatively few sibships.

REFERENCES

"Biometrika Tables for Statisticians" (1966). Pearson ES, Hartley HO (eds): Cambridge University Press.

Bishop DT, Cannings C (1978). Technical Report No. 11, Dept of Medical Biophysics and Computing, University of Utah.

Carmelli D, Karlin S, Williams RR (1978). A class of indices to assess major-gene versus polygenic inheritance of distributive traits (these proceedings).

Fain PR (1978). Characteristics of simple sibship variance tests for the detection of major loci, and application to height, weight, and spatial performance. Ann Hum Genet 42:109.

Fain PR, Ott J (1976). Heterogeneity of within sibship variance as a test of the major-gene hypothesis. Fifth International Congress of Human Genetics. Excerpta Med International Congress Series 397:180.

Mérat P (1968). Distributions des fréquences interprétation du determinisme génétique des charactères quantitatifs et recherche des "gènes majeurs." Biometrics 24:277.

Morton NE, MacLean CJ (1974). Analysis of family resemblance. III. Complex segregation of quantitative traits. Am J Hum Genet 26:489.

ACKNOWLEDGEMENT

This research was done while the authors were visiting the Department of Medical Biophysics & Computing, University of Utah, under NIH research grant CA-16573.