Types of Disease and Models for Their Genetic Analysis

by Robert C. Elston and Kadambari K. Namboodiri

Abstract

The authors compare schizophrenia with several other diseases and discuss how a few simple models that have already been successfully applied in other cases could be used in the genetic analysis of schizophrenia and MAO activity. Among the diseases discussed are Huntington's disease, xanthomatosis, and diabetes. The authors recommend undertaking multivariate studies of monoamine oxidase, dopamine β-hydroxylase, and other traits associated with schizophrenia in single, large pedigrees ascertained through schizophrenic probands.

For purposes of determining an appropriate method of genetic analysis, diseases may be characterized with respect to various criteria. They may be common or rare; have the same or different prevalences in the two sexes; be present at birth or develop only later in life; and they may be clearly diagnosable as homogeneous entities or subject to considerable clinical variability, and hence not always distinguishable from the "normal" state. Schizophrenia is a relatively common disease, equally prevalent in both sexes, with a variable age of onset; it is diagnosed on the basis of many symptoms none of which can always clearly distinguish individuals with and without the disease. Monoamine oxidase (MAO) activity, the main subject of this issue of the Bulletin, is measured as a simple quantitative trait. Although there can be considerable error in its determination, it is nevertheless a relatively objective measure. It can be studied in different tissues and with respect to several different substrates, and in this sense it is multivariate; often, however, it is reported as a univariate trait. We compare schizophrenia with several other diseases and discuss how a few simple models that have already been successfully applied in other cases could be used in the genetic analysis of schizophrenia and MAO activity.

Types of Disease

All diseases can be regarded as falling on a spectrum with respect to the relative importance of genetic and environmental factors in their pathogenesis. Toward one end of the spectrum, genetic factors are the most important; at the other end, environmental factors dominate. Essentially every disease is to some extent genetic and to some extent environmental. Close to the genetic end lie disorders like phenylketonuria and galactosemia; but even for these biochemical errors in metabolism, nongenetic factors can considerably affect the clinical characteristics or phenotype. At the other extreme, toward the environmental end, are infectious diseases like tuberculosis, where the causative agent can be identified; but individuals may differ genetically in their susceptibility to such diseases. In the middle of such a spectrum are disorders such as coronary heart disease, diabetes, schizophrenia, and affective disorder.

The rarer a disease is in the population, the easier it is to detect familial aggregation and, usually, the easier it is to determine etiology.

Reprint requests should be sent to Dr. Elston at Department of Biometry, Louisiana State University Medical Center, 1542 Tulane Ave., New Orleans, LA 70112.
Huntington’s disease, for example, is similar to schizophrenia in being equally prevalent in both sexes, having a variable age of onset, and not always being clearly diagnosable—in fact, patients with Huntington’s disease are frequently misdiagnosed as schizophrenics. Psychiatric disturbance is the first indication of the onset of Huntington’s disease in almost half the cases; there are often personality changes, hallucinations, and delusions as many as 25 years before any movement disturbances begin. The similarity between the two diseases may even extend to an underlying biochemical defect, since the only medications that provide relief for the choreic movements of Huntington’s disease are the routinely used to treat schizophrenia. But Huntington’s disease has less than one hundredth the prevalence of schizophrenia, and the familial aggregation is striking: There has been little difficulty in determining from extended pedigrees that in any one family Huntington’s disease is transmitted as a single autosomal dominant gene, with a penetrance that is incomplete only because some individuals die of other causes before their age of onset is due. Given that Huntington’s disease, with such a similar phenotype, is due to monogenic inheritance, we can hardly argue from the very nature of its phenotype that schizophrenia cannot be monogenic. However, we can expect that because of schizophrenia’s relatively high prevalence and less striking familial aggregation, it will be harder to identify any single gene underlying its etiology.

Xanthomatosis is perhaps not quite so rare as Huntington’s disease, and for a long time there was speculation that it was due to an incompletely dominant gene. However, it was not until a different trait—serum cholesterol level—was studied in the families where xanthomata were present that the picture began to become clearer (Harlan, Graham, and Estes 1966). In such families, cholesterol levels are bimodally distributed, the underlying dichotomy being transmitted as an autosomal dominant gene. Although the two component distributions overlap somewhat, it has been possible to confirm this mode of transmission by linkage analysis (Berg and Heiberg 1975; Elston et al. 1975). It should be noted that this simple mode of transmission is not clearly discernible if we restrict ourselves to studying either the phenotype “presence of xanthomata” or the phenotype “elevated cholesterol level” in the relatives of randomly selected probands with the corresponding phenotype: We need to study the phenotype “elevated cholesterol level” in the families of probands with xanthomata. If we are looking for a simple mode of transmission underlying schizophrenia, we should not necessarily study the phenotype “schizophrenia” (whatever that may be) in the families of randomly selected probands with this phenotype. If we close our eyes to all variability in these families other than whether or not each individual is classified as having (or having had) schizophrenia, we automatically make our goal more difficult. (Superficially, of course, the genetic analysis of a univariate all-or-none phenotype is simpler than the genetic analysis of a highly multivariate phenotype composed of both quantitative and qualitative traits. But our goal is not simply to perform a statistical analysis; it is to identify genes, if there are any, in the etiology of schizophrenia.)

Diabetes is a disease that is six or seven times more prevalent than schizophrenia, and its study is complicated by the following five factors, which similarly apply to the study of schizophrenia:

- The basic defect is unknown.
- There are no generally accepted criteria which allow the detection of prediabetes. Markers such as insulin response to glucose load or capillary basement membrane thickening have not proven to be universally accurate in the detection of prediabetes.
- Like schizophrenia, diabetes is extremely sensitive to environmental variables. Diet, obesity, viral infections, etc., confer increased risk toward the diabetic phenotype. Recently (Maugh 1979) a virus—Coxsackie B4—has been successfully isolated from a victim of juvenile onset diabetes (JOD). Epidemiological studies have shown both seasonal variation in the incidence of JOD and a temporal relation between certain viral infections and subsequent development of diabetes. Other viruses implicated are mumps virus and rubella virus.
- Similar to the geographic variability in schizophrenia, there is considerable ethnic variability in the prevalence of diabetes. Several North American Indian tribes have an unusually high prevalence of the disorder (Elston et al. 1974).
- There is clinical heterogeneity in diabetes. This heterogeneity implies that there are different genetic and/or environmental etiologic factors resulting in the various phenotypes.

This last point needs some elaboration in the case of diabetes. In recent years, critical analysis of clinical, physiological, biochemical, and genetic variability among diabetic patients has delineated several distinct
forms of diabetes (Rotter and Rimoin 1978). Patients with juvenile onset diabetes (JOD) are thin, ketosis-prone, and insulin-dependent; whereas those with maturity onset diabetes (MOD) are generally obese, nonketotic and insulin-resistant. These could be distinct entities, as are the several single gene mutations which cause abnormal glucose tolerance, obesity, and diabetic phenotype in different strains of mice (Coleman 1979). In a recent study of the erythrocyte insulin receptor sites of MOD patients and normals, there was a substantial (42 percent) reduction in the number of binding sites per cell in the patients as compared to controls. The deficit of insulin binding in MOD is thus associated with fewer binding sites per cell with little or no change in affinity to insulin (Robinson et al. 1979).

Evidence of genetic heterogeneity in diabetes comes from marker association studies, especially with the histocompatibility antigens. Studies of HLA associations show a definite excess of B8 and B15 antigens in JOD but not in MOD (Cudworth and Woodrow 1976; Ludwig, Scherthaner, and Mayr 1977). An even stronger association is seen between JOD and the HLA D locus (Rubenstein, Suciu-Foca, and Nicholson 1977), while such an association is not found with MOD. The most popular hypothesis for this association is that the HLA antigens serve as closely linked markers for diabetogenic genes, as yet untypable, on chromosome 6; other explanations of the phenomenon are, however, possible. Pyke and Nelson (1976) studied a large group of identical twins and found concordance for JOD to be about 50 percent, while for MOD it was nearly 100 percent. Concordant JOD pairs demonstrated an increased prevalence of HLA B8 compared to the discordant pairs. This further shows that different mechanisms are operating in these two major groups of diabetic patients. Another interesting feature is that autoimmune phenomena are prevalent in JOD but not in MOD. However, the Le(a\textsuperscript{-}b\textsuperscript{-}) phenotype was found in excess in both JOD and MOD by Vague et al. (1978).

Even within JOD, there is evidence for further heterogeneity (Rotter and Rimoin 1978), though this heterogeneity may not be genetic in origin. Tattersall and Fajans (1975) have been able to differentiate a distinct group of patients with early onset but MOD phenotype. Patients in this group have no ketonuria, and the disease can be controlled without insulin. Such maturity onset diabetes in youth (MODY) shows vertical transmission for three generations suggesting an autosomal dominant mode of inheritance with complete penetrance. This is in contrast to JOD families in which only 11 percent of the parents of JOD are diabetic. However, Spielman, Baker, and Zmijewski (1978) studied a group of JOD families with HLA information and report that a one-dose or dominant hypothesis is more compatible with their data than a recessive hypothesis.

We have stressed the fact that a disease may be heterogeneous, and discussed at some length how a recognition of this fact is being used in the genetic analysis of diabetes; but we must not be deluded into thinking that phenotypic heterogeneity necessarily implies genetic heterogeneity. It is generally recognized that in many single large pedigrees, von Willebrand's disease (vWD) follows a simple autosomal dominant mode of transmission, provided we assume that the carriers of this one gene are extremely heterogeneous in their phenotypic expression of it (Barrow and Graham 1964). Von Willebrand's disease is a rare hemorrhagic disorder with the following main differences from classic hemophilia: (1) prolonged bleeding time and (2) reduced factor VIII (antihemophilic factor, AHF) related activities such as factor VIII coagulant activity, factor VIII related antigen, and factor VIII related von Willebrand's factor. Clinical symptoms are characterized mostly by gastrointestinal, urinary, and uterine bleeding which may require repeated blood transfusions. Studies have shown lack of a clear relationship between laboratory test results and the degree of clinical expression of the disorder, making it difficult to identify carriers of the mutant allele who are asymptomatic. Recently, however, Goldin et al. (in press) have shown how a multivariate analysis can considerably reduce the misclassification errors; we shall return to this later to illustrate a multivariate model that could be used in the simultaneous analysis of schizophrenia and MAO.

Models for Genetic Analysis

We shall describe here two models suggested by Elston and Stewart (1971) for analyzing quantitative traits, showing how they can also be applied to (1) disorders with a variable age of onset, and (2) multivariate traits made up of both quantitative data and all-or-none clinical symptoms. We shall not go into the mathematical details of how the use of these models is implemented, a subject that has been reviewed elsewhere (Elston, in press; Elston and Rao 1978). The term "model" is used in the statistical sense of an underlying structure to the data, as-
assumed to be true whether or not there is a major locus segregating; under such a model, we test the null hypothesis of Mendelian segregation at a single locus.

In the first model it is assumed that there are three types of individuals, whom we label AA, Aa, or aa. The types of the individuals marrying into the families we study are determined in a probabilistic way by their prevalences in some population; but the types of the offspring are determined by the transmission of A from their parents. An offspring who receives A from both parents is type AA, one who receives A from just one parent is type Aa, and one who does not receive A from either parent is type aa. We define three transmission probabilities as follows: $\tau_{AA} = P$ (an AA individual transmits A to offspring); $\tau_{AA} = P$ (an Aa individual transmits A to offspring); $\tau_{AA} = P$ (an aa individual transmits A to offspring). It follows that under this general model the null hypothesis of Mendelian segregation is equivalent to the null hypothesis $\tau_{AA} = 1$, $\tau_{AA} = 1$, $\tau_{AA} = 0$, and the types of individual then represent genotypes. To analyze a quantitative trait the model incorporates, for each of the three types of individuals, a phenotypic distribution that is usually assumed to be Normal after an appropriate transformation of the data. These distributions can also be age- and/or sex-dependent, if necessary. This model has been used by Elston, Namboodiri, and Hames (1979) to analyze serum dopamine-$\beta$-hydroxylase (DBH) levels in families; after a logarithmic transformation, the data are found to be compatible with Mendelian transmission such that the phenotypic distributions for types AA and Aa are the same normal distribution and have a higher mean than the Normal phenotypic distribution for type aa—i.e., high levels of DBH could be transmitted as an autosomal dominant trait. Recently Demenais and Elston (in press) have shown that if this model is generalized to allow the three transmission probabilities to be dependent on the sexes of the transmitting parent and the offspring, this single model includes as null hypotheses not only autosomal transmission, but also X-linked transmission and some special cases of environmental, such as viral, transmission.

In the second model it is assumed that an individual's genotype is made up of a polygenic component and a major gene component; and, once again after suitable transformation, it is assumed that for each genotype there is a Normal phenotypic distribution. In this model it is assumed that the only transmission from one generation to the next is genetic. The model was extended by Morton and MacLean (1974) to incorporate in addition an environmental correlation among siblings. This model can distinguish between monogenic and polygenic inheritance, but only if we are certain that no other type of transmission from one generation to the next is involved; also, its use has proved to be very expensive of computer time so far.

With both these models, it is possible to analyze a disease with variable age of onset by substituting, instead of a Normal phenotypic distribution for a quantitative trait, a distribution that specifies, for each possible genotype and current age of an individual, the probability that the disease is expressed. Elston and Yelverton (1975) have proposed estimating two components of this "penetrance" probability: (1) the susceptibility, or probability that an individual will ever express the disease if he lives long enough; and (2) the age of onset distribution, the probability he will express the disease at (or by) a particular age if he is susceptible. In most sets of data there tends to be confounding between these two components, and so it is advisable to let only one of them be dependent on genotype when the penetrance is modeled in this way. A dominant gene such as for Huntington's disease, for example, can be modeled by assuming that all individuals have the same susceptibility while the age of onset distribution depends on genotype: for AA and Aa individuals it is Normal with a certain mean and variance, while for aa individuals it is Normal with a larger mean but the same variance. When this is done, it is found that the susceptibility is not significantly different from unity, and that the two age of onset distributions are as illustrated in figure 1: The mean and standard deviation of the age of onset distribution for AA and Aa individuals are 38.02 and 5.70, respectively; the analogus curve for aa individuals—which is hardly discernible because there are virtually no sporadic cases—is the left-hand tail of a Normal distribution with mean about 120 and the same standard deviation.

For mathematical simplicity it was assumed that conditional on the genotype at the major locus the ages of onset between any two relatives are uncorrelated; the results of the analysis did not appear to be materially affected by this perhaps unrealistic assumption. There is no difficulty in extending the model to allow for different disease classifications, such as mild and severe disease forms. More important for genetic analysis than severity, however, is the distinction between those forms of the disease that are suffi-
Figure 1. Cumulative age of onset distributions for individuals in families where Huntington's disease is segregating

The distribution estimated for aa individuals is barely discernible, having a mean of about 120 years and a standard deviation of 5.7 years.

Goldin et al. analyzed data on two unrelated large families, comprising 75 and 100 individuals, respectively, in which vWD is presumed to be segregating. In addition to the presence or absence of various clinical symptoms, quantitative laboratory data on factor VIII activities and bleeding time are available on each person. As most of the symptom variables are highly correlated, the first step is to reduce them to a single principal component score (Morrison 1976) designated the symptom index. In this way the set of 0.1 variables (i.e. absence or presence of each symptom) is converted into a quantitative trait. This symptom index and the other quantitative traits are then transformed to reduce the skewness in them, and adjusted for sex and age effects. We then ask the question: What linear function (or weighted average) of these quantitative traits is most likely to be segregating as a single Mendelian autosomal gene? This question is answered by obtaining the maximum likelihood estimates of the coefficients in the linear function, on the assumption that the variability of this function in these pedigrees is, as much as possible, due to segregation at a single locus. This is analogous to a factor analysis in which there is a single factor hypothesized, except that we require that the factor represent what Mendel originally called a factor, i.e., a monogenically segregating mechanism. Goldin et al. performed this analysis in each family separately, and found that the factor representing a dominant autosomal gene was similar in the two families; in fact the factor score estimated in one family could be used to define a dominantly segregating quantitative trait in the other family with only a small overlap between the phenotypic distributions of the carriers and noncarriers of the mutant allele. Singly, however, each of the univariate traits showed virtually no evidence of major gene segregation.

Conclusion

There is no reason to believe that the genetic analysis of schizophrenia and MAO will require the use of models that are basically different from the ones we are already using in the genetic analysis of other traits. Their application, on the other hand, will need to be specific to the problem in hand, and any neurobiological or psychological insights that can be brought to bear on the problem should be incorporated. We know that schizophrenia is a highly multivariate trait, and there is now no doubt that MAO levels tend to be low in schizophrenics (Murphy and
Kalin 1980; this issue); it has also been reported that DBH levels are on an average lower in schizophrenics (Böök, Wetterberg, and Modrzska 1978). The fact that MAO levels appear to be lowered only in the subgroup of schizophrenics with well-defined auditory hallucinations and delusions (Schildkraut et al. 1980; this issue) suggests that schizophrenia, like diabetes, is genetically heterogeneous. The hemoglobinopathies provide a model example of genetic heterogeneity, the elucidation of which was largely accomplished by the careful study of single families. There is every reason to believe that a multivariate study of MAO, DBH, and other traits associated with schizophrenia, in single large pedigrees ascertained through schizophrenic probands, will eventually yield Mendelian "factors"; and the study of large pedigrees will have the additional advantage of making linkage analysis, to confirm such genes, more powerful.

Until we have discovered an unequivocal biochemical or neuropsychophysiological marker for schizophrenia, the computation of a linear function score incorporating all the multivariate data would serve the purpose of identifying carriers of a mutant allele with high accuracy. If such a trait is found to segregate in families and is linked to some polymorphic marker, the presence of a major locus for it would be confirmed, and the precision of detecting genotypes early in life before any clinical symptoms appear would be enhanced.

References


Berg, K., and Heiberg, A. Linkage studies on familial hyperlipo-


Schildkraut, J.J.; Orsulak, P.J.; Schatzberg, A.F.; and Herzog, J.M. Platelet monoamine oxidase activity...


**Acknowledgment**

The research reported was supported, in part, by a Public Health Service Research Scientist Award (MH 31732) and a research grant (MH 26721) from the National Institute of Mental Health, and a research grant (GM 16697) from the National Institute of General Medical Sciences.

**The Authors**

Robert C. Elston, Ph.D., is Professor and Head, Department of Biometry, Louisiana State University Medical Center, New Orleans, LA. Kadambari K. Namboodiri, Ph.D., is Research Associate Professor, Department of Biostatistics and the Genetics Curriculum, University of North Carolina at Chapel Hill, Chapel Hill, NC.