

Analyzing the Relationship Between Age at Onset and Risk to Relatives

Michael C. Neale,* Lindon J. Eaves,**† John K. Hewitt,* Charles J. MacLean,†
Joanne M. Meyer,* and Kenneth S. Kendler**†

Departments of *Human Genetics and †Psychiatry, Medical College of Virginia, Richmond

Summary

Correlations in age at onset between relatives affect risk to relatives of a given age. Either an increase or a decrease in risk may be observed for a relative of a proband, according to whether there is a causal relationship between liability to disease and age at onset. Likelihood formulas are given for pairs of relatives under a number of different sampling schemes, and it is shown how data collected from relatives enable maximum-likelihood estimation of parameters of a linear model relating disease liability and age at onset. A genotype-environment extension of this model was fitted to data on age at onset for schizophrenia that were obtained from the National Academy of Sciences–National Research Council Twin Registry. Age at onset is correlated between twins, but this correlation appears to be associated with factors that are separate from those which affect liability to disease. However, even this relatively large sample of twins is too small to draw firm conclusions about any causal relationship between disease liability and onset.

Introduction

Age at onset is variable for many physical and mental diseases, including hypercholesterolemia and heart disease (Heilberg and Slack 1977; Rissanen 1979), neurodegenerative disease (Ridley et al. 1986), breast cancer (Anderson 1972), schizophrenia (Kendler et al. 1987), Alzheimer disease (Sturt 1986), Huntington chorea (Pericak-Vance et al. 1983), and major depression (Price et al. 1987). Information about age at onset is often gathered from samples of patients and used to assess the risk to relatives. Weinberg (1925) appears to be the first to have addressed the problem, and subsequently Strömberg (1935, 1938) proposed two different methods of adjusting morbidity risks. These methods were summarized by Larsson and Sjögren (1954), and, more recently, a maximum-likelihood approach was described by Risch (1983). The early methods rely

on a weighting procedure to adjust estimates of risk in relatives according to a function of the elapsed period of risk. Risch allows for the simultaneous estimation of an age-at-onset function. However, none of the methods addresses the possibility that age at onset may be an index of liability to disease.

In the present article, we show that a correlation between liability to disease and age at onset can lead to quite different morbidity risks for relatives of an affected individual with a given age at onset. We describe a model that allows for age-at-onset correlation to be caused by (1) resemblance in liability, which determines, in part, age at onset, and (2) resemblance for factors that affect age at onset but that are independent of disease liability. Maximum-likelihood estimation of the parameters of this model is possible using data collected from pairs of relatives under a variety of ascertainment schemes. The model is extended to partition genetic and environmental sources of variance and is applied to data on schizophrenia from the National Academy of Sciences–National Research Council (NAS-NRC) Twin Registry. The approach is suitable for any potential index of liability that can be measured only in affected individuals, such as response to treatment or certain psychological or physiological tests.

Received December 15, 1988; final revision received April 14, 1989.

Address for correspondence and reprints: Michael C. Neale, Department of Human Genetics, Medical College of Virginia, Box 33, Richmond, VA 23298.

© 1989 by The American Society of Human Genetics. All rights reserved. 0002-9297/89/4502-0005\$02.00

Model for Liability and Age at Onset

For simplicity, we assume a normal distribution of liability to disease, with an abrupt threshold t . Individuals with liability values above this threshold will become affected if they complete the age of risk. This is the model described by Pearson (1900) and more recently by Wright (1934) and Falconer (1965); it is currently popular in both psychiatric genetics (e.g., see Cloninger et al. 1978; McGue et al. 1983) and behavioral genetics (e.g., see Neale et al. 1986; Heath et al. 1985, and in press; Neale 1988). We allow the liability to disease to correlate between relatives. Our normality assumption is one of convenience rather than necessity; alternative forms for the liability distribution, such as may be obtained from a major-locus model or a mixed model (Elston and Stewart 1971) may be selected. In principle, the moments of the distribution of age at onset provide information to discriminate between these models, but it seems likely that statistical power for this comparison is very low (e.g., see Eaves 1983).

The same normality assumption is made with respect to variation on a scale of age at onset. Theoretically, all individuals have an age-at-onset value, but this can be observed only in persons above the threshold of liability. We assume a simple linear model of the relationship between liability to disease and age at onset. Two other causes of age at onset are specified: a component C which is shared by family members (owing to genetic and environmental covariation) and a residual random environment or error component (D_i). Both causes are independent of liability to disease. A path diagram of this model of familial resemblance is shown in figure 1. Although the exogenous variables L_1 , L_2 , C , D_1 , and D_2 are assumed to be normally distributed, truncation of the liability variables will cause nonnormality of the observed ages at onset in the presence of a nonzero value of the path b from liability to age at onset. In general, this nonnormality will be minor unless the value of b is very large (say, $>.9$). Nevertheless, the effects of truncation on the correlation between variables may be substantial (e.g., see Fisher 1931; Aitken 1934; Curnow 1972; Johnson and Kotz 1972; Smith and Mendell 1974; Curnow and Smith 1975; Bucher and Schrott 1982; Martin and Wilson 1982; Neale et al. 1989).

Recently, there have been several efforts to develop methods for the analysis of data from truncated samples. Boehnke and Lange (1984) specify a likelihood function conditional on the proband's actual pheno-

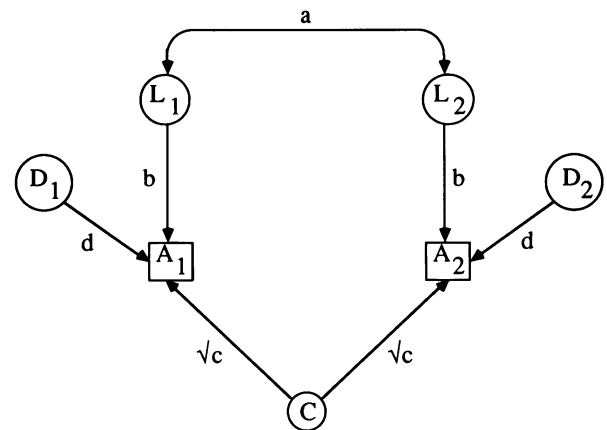


Figure 1 Causal model of familial resemblance for liability (L_i) and age at onset (A_i). C = genetic and common environmental age-at-onset factors that are independent of disease liability; D_i = residual variation on age at onset, including measurement error.

typic value, to permit the analysis of data collected from “enriched” samples in which probands are diagnosed if their position on a continuous scale exceeds a threshold. Rao et al. (1988) employ a constraint on the mean of probands to obtain parameter estimates with smaller standard errors under the same “direct” truncation model. Rao and Wette (1987) compare direct truncation with “indirect” truncation, which is made through a correlated phenotype, and with “latent truncation,” which is made through a latent variable, e.g., liability to disease. In the present article, latent truncation is assumed; our treatment necessarily differs, since age at onset of disease cannot be observed in unaffected individuals.

Likelihood Formulas

Initially, we consider age at onset and liability to disease to be independent of the age structure of the population. Dependency may occur if there is either increased mortality associated with liability to disease or (in the case of parents and offspring) reduced reproductive fitness following disease onset (Heimbuch et al. 1980).

Samples Without Relatives

To clarify the treatment for pairs of relatives, we confine our initial discussion to samples drawn from the population, without regard to family structure. The formulation is described in general terms for the following three basic schemes of ascertainment: (1) random sampling of subjects of a given age, (2) random

sampling irrespective of age, and (3) sampling within a particular age band.

We define terms as follows: q = liability; r = age-at-onset value; s = current age; t = threshold on liability distribution above which individuals will become affected; $\phi(x)$ = multinormal density function; and $\psi(z)$ = the age distribution in the population. The age-at-onset value is a latent variable in this treatment; individuals may have a predisposition to early age of onset but may not show such onset because they are below the liability threshold.

The relationship between age at onset and disease liability of an individual may be considered as a two-way contingency table. A graphical representation of this table is shown in figure 2, for a correlation of $-.8$ between age at onset and liability. Five distinct regions, R1–R5, are shown. Only those individuals in region R4 are affected, having both liability above threshold and onset prior to current age. Those in region R2 have sufficient disease liability but have not yet reached age

at onset, while those in regions R1 and R3 (cutaway portion of fig. 2) are below threshold. The plane R5 gives the proportion of individuals above threshold for a particular value of the age-at-onset distribution (r), in this case the current age (s). We note that for computational purposes the region R1 may be integrated with either R2 or R3, thus saving considerable computer time.

Subjects with a particular age.—If a sample is drawn such that all subjects have the same age, s , the likelihood of observing an affected individual may be written as a two-dimensional integral:

$$L(+|s) = \int_t^\infty \int_{-\infty}^s \phi(q,r) drdq, \tag{1}$$

which, to simplify notation, we write as

$$L(+|s) = \int_{R4} \phi(\underline{x}) d\underline{x}, \tag{2}$$

where $\underline{x} = (q,r)$, i.e., liability and age at onset, s is

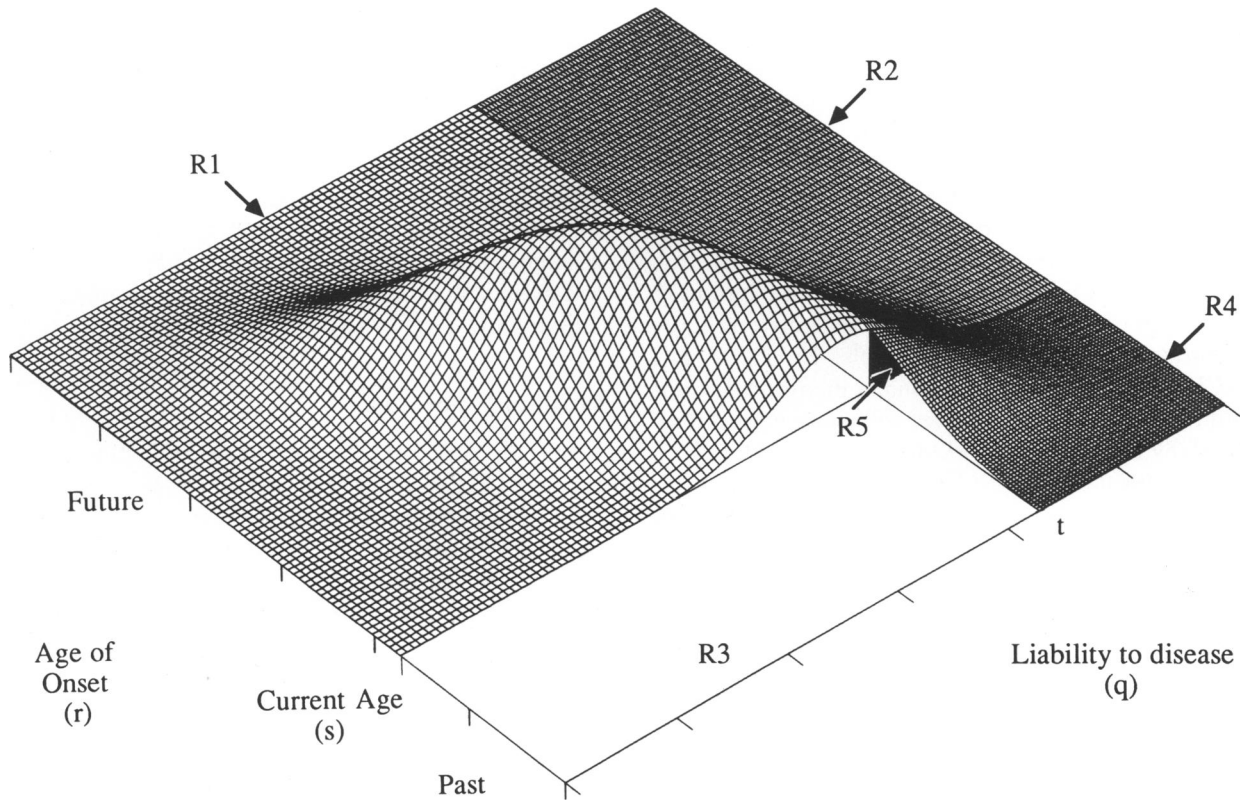


Figure 2 Bivariate normal distribution with $.8$ correlation between disease liability (q) and age at onset (r). Four regions (R1–R4) are defined according to whether an individual is above or below liability threshold and according to whether age at onset is before or after current age. The plane R5 defines the probability that an individual is above threshold for a particular value of r , in this case where $r = s$.

current age, and $R4$ is the region shown in figure 2. In the same notation, the likelihood for unaffected individuals is

$$L(-|s) = \sum_{i=1}^3 \int_{R_i} \phi(\underline{x}) d\underline{x}, \quad (3)$$

being the sum of the three two-dimensional integrals described by regions $R1$, $R2$, and $R3$ shown in figure 2.

To describe the likelihood for an individual aged s with a particular age at onset, r , we define $R5$ as the area where liability is above threshold and age at onset is r . This region is shown in figure 2 for the case $r = s$. The likelihood is therefore

$$L(+, r, s) = \int_{R5} \phi(\underline{x}, \underline{y}) d\underline{x}, \quad (4)$$

where $\underline{x} = (q)$ and $\underline{y} = (r)$; we note that the bivariate normal density of $(\underline{x}, \underline{y})$ is integrated over liability \underline{x} but not over the observed age at onset, \underline{y} .

Random sampling.—Many samples do not consist of subjects of a particular age, so we now consider random sampling from a population in which the age distribution is given by $\psi(z)$. If the age distribution and liability are independent, the likelihood under random sampling (L_R) that an individual is affected ($q > t$) is

$$L_R(+, s) = \int_{R4} \int_0^\infty \phi(\underline{x}) \psi(z) dz d\underline{x}, \quad (5)$$

where $\underline{x} = (q, r)$, the liability and age-at-onset variables, $z = (s)$, a dummy age variable, and $R4$ is the region corresponding to those individuals above liability threshold with age at onset less than current age, as shown in figure 2.

In practice, it is often possible to establish an age, u , above which the probability of onset is effectively zero. If a sample consists entirely of individuals with age greater than u , the likelihood simplifies to the one-dimensional integral

$$L_R(+, s) = \int_{R5} \phi(\underline{x}) d\underline{x} \quad \text{if } s > u, \quad (6)$$

where $\underline{x} = (q)$, the liability. The likelihood of observing an unaffected individual, $L_R(-, s)$, is simply $1 - L_R(+, s)$.

To observe an individual with a particular age at onset, r , they must have age $\geq r$, with onset $y = r$ and liability $q > t$, so the likelihood is

$$L_R(+, r) = \int_{R5} \int_r^\infty \phi(\underline{x}, \underline{y}) \psi(z) dz d\underline{x}, \quad (7)$$

where $\underline{x} = (q)$, the liability; $\underline{y} = (r)$, the age at onset;

and $z = (s)$, the dummy age variable. If the sample consists entirely of individuals past the age at risk ($s > u$), we have

$$L_R(+, r) = \int_{R5} \phi(\underline{x}, \underline{y}) d\underline{x} \quad \text{if } s > u, \quad (8)$$

where $\underline{x} = (q)$ and $\underline{y} = (r)$ as in equation (7). Hence, the likelihood in this case is expressed as a one-dimensional integral.

Sampling in a particular age band.—Suppose ascertainment is through affected individuals in an age band (v_1, v_2) . The proportion of individuals in the population that will be ascertained (i.e., the ascertainment correction) is

$$P = \int_{R4} \int_{v_1}^{v_2} \phi(\underline{x}) \psi(z) dz d\underline{x}, \quad (9)$$

where $\underline{x} = (q, r)$, liability and age of onset; $z = (s)$, the dummy age variable; and $R4$ is affected persons as defined in figure 2. Under this banded sampling, the likelihood (L_B) of observing, in an ascertained individual, a particular age at onset that is less than the current age ($r < s$) is

$$L_B(+, r) = \frac{1}{P} \int_{R5} \int_{v_1}^{v_2} \phi(\underline{x}) \psi(z) dz d\underline{x}, \quad (10)$$

where $\underline{x} = (q, r)$, liability and age at onset; and $R5$ is the proportion of individuals above threshold, as defined in figure 2. The likelihood of observing an individual of a particular age, s , with a particular age at onset, r , is

$$L_B(+, r, s) = \frac{1}{P} \int_{R5} \phi(\underline{x}, \underline{y}) \psi(z) d\underline{x}, \quad (11)$$

where $\underline{x} = (q, r)$, liability and age at onset; and $R5$ is the proportion of individuals above threshold, as defined in figure 2. The case of ascertainment without regard to age may be obtained by setting $v_1 = -\infty$ and $v_2 = \infty$.

If age at onset and liability are not independent of the age distribution (owing to the effects of mortality or fertility), then we may substitute $\psi(z|\underline{x}, \underline{y})$ for $\psi(z)$ throughout the above expressions. However, the specification of $\psi(z|\underline{x}, \underline{y})$ may not be simple.

Pairs of Relatives

Regardless of sampling scheme, data from unrelated individuals do not provide the information to estimate the parameters of the model in figure 1. Therefore we generalize the likelihood formulations given above to pairs of relatives. Four basic schemes of sampling are

considered: (1) random samples of individuals of a given age, (2) affected individuals and their relatives at a given age, (3) affected individuals and their relatives in an age band, and (4) pairs past the age of risk. The complexity of the likelihood expressions—and the consequent feasibility of analysis—differs considerably under these sampling strategies.

Samples of a given age.—Suppose that pairs of relatives are obtained as a result of population screening at a particular age, s . The likelihood of concordant pairs with ages at onset r_1 and r_2 is the two-dimensional integral

$$L(+, +, r_1, r_2 | s) = \int_{R_1} \int_{R_2} \phi(\underline{x}, \underline{y}) d\underline{x}, \quad (12)$$

where $\underline{x} = (q_1, q_2)$, the liabilities; and $\underline{y} = (r_1, r_2)$, the ages at onset. This likelihood is the bivariate extension of equation (4) above; it varies according to the covariance between relatives. For discordant pairs, the likelihood is

$$L(+, -, r_1 | s) = \sum_{i=1}^3 \int_{R_5} \int_{R_i} \phi(\underline{x}, \underline{y}) d\underline{x}, \quad (13)$$

where $\underline{x} = (q_1, q_2)$, the liabilities; and $\underline{y} = (r_1)$, the age at onset of the affected relative. The likelihood for pairs concordant for being normal is

$$L(-, -, |s) = \sum_{i=1}^3 \sum_{j=1}^3 \left\{ \int_{R_i} \int_{R_j} \phi(\underline{x}) d\underline{x} \right\}, \quad (14)$$

where $\underline{x} = (q_1, q_2, r_1, r_2)$. The nine terms in this expression reduce to four when R_1 and R_2 are considered jointly. The four terms correspond to the following cases: (1) both relatives are below threshold; (2) relative 1 is above threshold but has not reached his age at onset, and relative 2 is below threshold; (3) relative 2 is above threshold but has not reached his age at onset, and relative 1 is below threshold; and (4) neither relative has reached age at onset, though both are above threshold. The likelihood for concordant normal pairs of a given age is invariant over pairs. Hence, computation of the four-dimensional integrals (eq. [14]) is required only once for any set of parameter values.

Ascertainment through affected relatives.—If pairs of affected relatives are ascertained because at least one member of each pair is sick, the four-dimensional integrals for concordant normal pairs seem to be avoided. However, it is necessary to introduce a correction for ascertainment that involves precisely the same four-dimensional integral, equal to $1 - L(-, -, |s)$. For disorders with a low cumulative lifetime incidence, a

random sample may prove inefficient, as most data collection will be from concordant unaffected pairs. These pairs will be relatively uninformative if, like age at onset, the index can only be measured in affected individuals. Nevertheless, the random sample has the advantage that the population threshold may be estimated from the data, rather than supplied as a fixed parameter.

Sampling affected relatives in an age band.—The ascertainment correction is more complex if we consider sampling within a particular age band (v_1, v_2). This situation is shown graphically in figure 3. If we require both relatives to be within the age band, then only area Q1 is sampled. We write the two-dimensional integral that describes this area as

$$\int_{Q1} \psi(\underline{s}) d\underline{s}, \quad (15)$$

where $\underline{s} = (s_1, s_2)$, the dummy age variables. If pairs are ascertained if at least one member of the pair is affected, and if both members are in the age range, we have the correction

$$P_2 = \int_{Q1} \psi(\underline{s}) d\underline{s} - \sum_{i=1}^3 \sum_{j=1}^3 \int_{R_i} \int_{R_j} \int_{Q1} \phi(\underline{x}) \psi(\underline{s}) d\underline{s} d\underline{x}, \quad (16)$$

where $\underline{x} = (q_1, q_2, r_1, r_2)$, liabilities and ages at onset; and $\underline{s} = (s_1, s_2)$, the dummy age variables. The likelihood for concordant pairs is

$$L(+, +, r_1, r_2, s_1, s_2) = \frac{1}{P_2} \int_{R_4} \int_{R_4} \phi(\underline{x}, \underline{y}) \psi(\underline{s}) d\underline{x}, \quad (17)$$

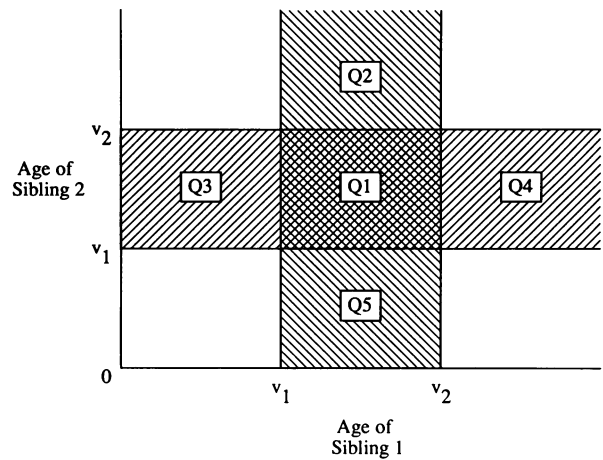


Figure 3 Diagram showing regions Q1–Q5, which describe area in which either one (Q2–Q5) or both (Q1) siblings lie within a sampling age band (v_1, v_2).

where $\underline{x} = (q_1, q_2)$, the liabilities; $\underline{y} = (r_1, r_2)$, the ages at onset; and $\underline{s} = (s_1, s_2)$, the observed current ages. For discordant pairs, we have

$$L(+, -, r_1, r_2, s_1, s_2) = \sum_{i=1}^3 \frac{1}{P_2} \int_{R_4} \int_{R_i} \phi(\underline{x}, \underline{y}) \psi(\underline{s}) d\underline{x}, \quad (18)$$

where $\underline{x} = (q_1, q_2, r_2)$, the liabilities and age of onset in the normal relative; $\underline{y} = (r_1)$, the observed age at onset in the affected relative; and $\underline{s} = (s_1, s_2)$, the current ages. Hence, although the maximum number of dimensions of integration is three for any pair of observations in the data, the ascertainment correction involves four six-dimensional integrals, computation of which would be prohibitive with current computer resources and software.

Even more complex is the case where pairs are ascertained if one affected member is in the age range but the relative is ascertained irrespective of age. In this case, the ascertainment correction consists of the set of pairs that could be sampled by meeting the age criteria, but not concordant for normal status:

$$P_1 = \sum_{k=1}^5 \left[\int_{Q_k} \psi(\underline{s}) d\underline{s} - \sum_{i=1}^3 \sum_{j=1}^3 \int_{R_i} \int_{R_j} \int_{Q_k} \phi(\underline{x}) \psi(\underline{s}) d\underline{s} d\underline{x} \right], \quad (19)$$

where \underline{x} and \underline{s} are as defined in equation (16). The likelihoods for concordant and discordant pairs are the same as in equations (17) and (18) above, but with denominator P_1 .

Samples of pairs past the age of risk.—In practice, given that we are not prepared to assume that disease liability and age at onset are independent, we may be limited to the use of data that have been collected from individuals who have passed the age of risk. The likelihood for concordant pairs reduces from the general case, described above, to

$$L(+, +, r_1, r_2) = \frac{\int_{R_5} \int_{R_5} \phi(\underline{x}, \underline{y}) d\underline{x}}{1 - \int_{R_6} \int_{R_6} \phi(\underline{x}) d\underline{x}}, \quad (20)$$

where $\underline{x} = (q_1, q_2)$, $\underline{y} = (r_1, r_2)$, and R_6 is the region below liability threshold for a given age-at-onset value ($\int_{R_5} + \int_{R_6} = 1$). The likelihood for discordant pairs is

$$L(+, -, r_1) = \frac{\int_{R_5} \int_{R_6} \phi(\underline{x}, \underline{y}) d\underline{x}}{1 - \int_{R_6} \int_{R_6} \phi(\underline{x}) d\underline{x}}, \quad (21)$$

where $\underline{x} = (q_1, q_2)$ and $\underline{y} = (r_1)$. This formulation is subject to bias if mortality and either liability or age at onset are not independent.

Calculation of Risk to Relatives

Suppose a subject of known age presents for genetic counseling with an affected sibling, whose age at onset is known. We wish to calculate the probability that this counselee will become affected, given that he or she is currently of normal status. The counselee could be a member of one of two mutually exclusive classes: (1) above the threshold but has not reached age at onset (region R_2 in fig. 2) or (2) below the threshold and will never contract the disease (regions R_1 and R_3 in fig. 2).

Let subscripts 1 and 2 refer, respectively, to the affected and normal individuals in our pair of relatives. The probability that both siblings are above the threshold and that sibling 1 has onset r_1 is given by

$$\text{Prob}_2 = \int_{R_5} \int_{R_4} \phi(\underline{x}, \underline{y}) d\underline{x}, \quad (22)$$

where $\underline{x} = (q_1, q_2, r_2)$, the liabilities and age at onset in the counselee; and $\underline{y} = (r_1)$, the observed age at onset in the relative. Similarly, the probability that sibling 2 is below the threshold is

$$\text{Prob}_1 = \sum_{i=1}^2 \int_{R_5} \int_{R_i} \phi(\underline{x}, \underline{y}) d\underline{x}. \quad (23)$$

The risk to the individual being counseled is thus

$$\frac{\text{Prob}_2}{\text{Prob}_1 + \text{Prob}_2} \quad (24)$$

Differences in risk across the lifespan result from changes in the probability that the individual is above threshold, given that he or she is currently not affected. Curnow (1974) provides tables of relative risk for several different parameter values for the model shown in figure 1. However, he addresses the measurement of indices in general; our treatment differs because we make use of the additional information that the counselee is healthy at his or her present age.

The conditional risks were calculated using a FORTRAN program which uses subroutines D01BBF and D01FBF from the Numerical Algorithms Group (NAG) library (NAG 1988). Calculation of integrals with limits $-\infty$ and ∞ was performed using Gauss-Hermite quadrature, and Gauss-Laguerre quadrature was employed for integrals with finite lower limits. Thirty-two abscissas were used for each dimension of integration.

Table 1

Risk to a Healthy Individual for a Disorder with 1% Cumulative Lifetime Incidence and Familial Resemblance for Liability, Given His or Her Current Age and the Age at Onset in His or Her Relative

AGE OF COUNSELEE	$a = .7; b = 0; c = 0$ Relative's Age at Onset					$a = .7; b = 0; c = .448$ Relative's Age at Onset					$a = .7; b = -.8; c = 0$ Relative's Age at Onset				
	-2.0	-1.0	.0	1.0	2.0	-2.0	-1.0	.0	1.0	2.0	-2.0	-1.0	.0	1.0	2.0
-2.0	.26	.26	.26	.26	.26	.24	.26	.26	.27	.27	.37	.29	.25	.22	.21
-1.0	.23	.23	.23	.23	.23	.17	.21	.24	.26	.26	.32	.25	.22	.20	.18
.0	.15	.15	.15	.15	.15	.05	.10	.15	.20	.23	.20	.16	.14	.12	.11
1.0	.05	.05	.05	.05	.05	.01	.02	.05	.09	.14	.07	.05	.04	.04	.04
2.0	.01	.01	.01	.01	.01	.00	.00	.01	.01	.04	.01	.01	.01	.01	.00

NOTE.—Ages are standardized to that age-at-onset distribution. Model parameters (see fig. 1) correspond to three simple models: (1) correlation in liability to disease alone; (2) correlation between disease liabilities, and correlation between ages at onset; and (3) correlation in liability to disease that is a cause of age at onset. In both cases (2) and (3), model parameters yield a correlation of .448 between ages at onset in relatives.

To illustrate the effects that different sources of familial resemblance for age at onset have on the risk to relatives, we selected three sets of parameter values for the model shown in figure 1. All models were applied using a 1% cumulative lifetime incidence of the disorder in the population, with a correlation in disease liability of .7 between relatives. This correlation is high for siblings but is not uncommon for monozygotic (MZ) twin resemblance. The high value serves to emphasize the pattern of risks which would be observed for smaller correlations. Table 1 shows risk figures for five different standardized ages at onset in the relative and five different current ages of the counselee, for each of three special submodels. The first model specifies (1) no causal relationship between liability and age at onset and (2) no residual familial resemblance for age at onset. Here, the relative risks to the counselee depend only on the counselee's own age and are unaffected by the age at onset in the relative. This simple picture is implicit in many of the current approaches to age correction of risk.

The second set of parameter values specifies correlation in age at onset to be produced entirely by factors independent of liability. Under this model, we observe a general pattern of increasing risk with increasing age at onset in the relative. When the relative has early age at onset, the risk to the counselee drops more sharply than for the case when age at onset is not correlated between relatives. However, when the relative has late onset, the reverse is true, with the risk declining more slowly as the counselee ages. These consequences fol-

low logically from the expectation that relatives will have similar ages at onset, but age at onset does not provide any information about the counselee's liability to disease.

Our third example is arranged to give the same correlation in age at onset for pairs of relatives as the second example, but the resemblance is entirely due to a strong negative correlation between age at onset and disease liability ($b = -.8$). Under this model the risk to the counselee declines with increasing age at onset in the relative. Compared with the first example, the risk to the counselee is uniformly higher if the relative had an early age at onset and is lower if the relative had a late age at onset. This results from the information about the position of the counselee on the liability scale, information that is derived from the measurement of the correlated age at onset in the relative. Clearly, estimation of the population correlation matrix between the liability and age variables is essential for accurate prediction of risk to relatives and for understanding of the biology of the disease process.

Parameter Estimation

The estimation of parameters describing the relationship of age at onset to liability to disease depends on calculation of the likelihood of data collected from pairs of relatives. Each pedigree will normally have a different likelihood, so the computational demands increase with sample size. During parameter estimation, the likelihoods must be calculated for each set of param-

ter values used during the search. It is necessary to use numerical methods to integrate the multinormal distribution; the time taken to achieve a given level of accuracy may be expected to increase by a factor of 16 for each additional dimension of integration. With hardware such as a VAX 8650 and when general integration routines supplied by NAG are used, the estimation of parameters for likelihoods involving two dimensions is feasible for large sample sizes, taking approximately 3 h of CPU time for a sample size of 500. A maximum of three dimensions seems feasible at present.

Parameter estimation for data collected from pairs of relatives who have passed the age of risk is performed in a FORTRAN program which uses both a combination of adaptive and Gaussian routines in the NAG library (D01BBF, D01FBF, and D01EAF) for quadrature and a general routine for minimization subject to nonlinear parameter constraints (E04UCF).

Simulation

To test the program and to obtain some sense of the precision of parameter estimates, we employed Monte Carlo methods. For each of a range of parameter values, data were simulated from a truncated multinormal distribution by using the procedure described elsewhere by two of the authors of the present paper (Hewitt and Neale, in press). Five hundred pairs of relatives were

simulated for a variety of parameter values of the model. Parameters *a*, *b*, and *c* were set at .0, .4, or .8, and the population mean and variance were set at zero and unity, respectively. For a disease frequency of 10%, results of fitting the model to the simulated data are shown in table 2. For less frequent cases, it would be necessary to increase the sample size or to restrict simulation to cases in which there is a substantial correlation between relatives; otherwise the number of concordant pairs would be too small to provide meaningful results. Because the model is standardized, i.e., all variables have unit variance, no results are given for cases in which both *b* and *c* were set at .8, as this yields a variance greater than unity for the age-at-onset variable.

The simulation results are encouraging, but there are some discrepant values. The same solutions were obtained for a variety of starting values of parameter estimates, so local minima do not seem to be a problem. The parameter *c* seems especially liable to error. The information for this statistic is exclusively obtained from the concordant pairs, and these are rare when the correlation in liability is low. Sample sizes of 500 pairs would seem to be small for estimating the association between an age at onset and liability to disease with a cumulative lifetime incidence of 10%. However, the information obtained in this way can be of considerable importance, and the properties of efficiency and asymptotic absence of bias of maximum-likelihood estimators

Table 2
Estimates \hat{a} , \hat{b} , and \hat{c} for Simulated Data of 500 Pairs of Relatives: Proband Selection with Threshold of 1.282 (10% selection)

<i>c</i> _{true}	<i>a</i> _{true}								
	.0			.4			.8		
	<i>b</i> _{true}			<i>b</i> _{true}			<i>b</i> _{true}		
	.0	.4	.8	.0	.4	.8	.0	.4	.8
.0:									
\hat{a}	-.70	.00	-.03	.39	.34	.40	.79	.80	.83
\hat{b}02	.41	.80	-.01	.41	.81	-.03	.41	.80
\hat{c}00	.00	.00	.00	.00	.00	.00	.00	.00
.4:									
\hat{a}	-.05	-.03	-.01	.48	.42	.40	.80	.79	.78
\hat{b}	-.01	.42	.80	-.01	.39	.80	.01	.40	.79
\hat{c}00	.49	.54	.69	.00	.33	.48	.40	.50
.8:									
\hat{a}00	-.0841	.3680	.80	...
\hat{b}01	.38	...	-.02	.40	...	-.01	.41	...
\hat{c}00	.0091	.9283	.80	...

(Fisher 1922) indicate that the method presented here is optimal for assessing the importance of indexes of liability to disease.

Partitioning Familial Resemblance

In the previous section, the resemblance between relatives is simply characterized as a correlation, for both liability and the independent age-at-onset parameters. When data are collected from genetically informative groups, such as MZ and dizygotic (DZ) twins or adopted and biological relatives, familial resemblance may be partitioned into genetic and environmental components. In this section, the model in figure 2 is extended for use with twin data.

Under a simple additive linear model for genetic and environmental variation, we may write the structural equation $L = h_L G_L + e_L E_L + c_L C_L$, where L is the latent liability to disease, G_L is the additive genetic component of variation in liability, and E_L and C_L are the individual unique and common environmental components of variation, respectively. In data from twins reared together, the effects of genetical dominance and common environmental variance are confounded (Eaves 1970; Jinks and Fulker 1970). The sample sizes required to detect dominant genetic variation in liability are large even for variables measured on a continuous scale (Martin et al. 1978). We assume the effects of genetic dominance to be zero in our model; if this assumption were false, the proportion of common environment variance would be underestimated. Similarly, variation in age at onset may be partitioned into genetic and environmental components, together with the effect of liability on age at onset, giving $A = h_A G_A + e_A E_A + c_A C_A + bL$, where A is the latent liability to disease, and G_A , E_A , and C_A are, respectively, the additive genetic, individual unique, and common environmental components of variation in factors specific to age at onset. This model is shown in figure 4 as a path diagram.

NAS-NRC Schizophrenia Data

Subjects and Methods

The NAS-NRC Twin Registry consists of 15,924 pairs of Caucasian male twins born during the years 1917–27, all of whom served in the U.S. armed forces. The sample has been described in detail elsewhere (Jablon et al. 1967; Kendler and Robinette 1983; Kendler et al. 1987). Zygosity diagnosis was made on the basis of questionnaire response in most cases and was found

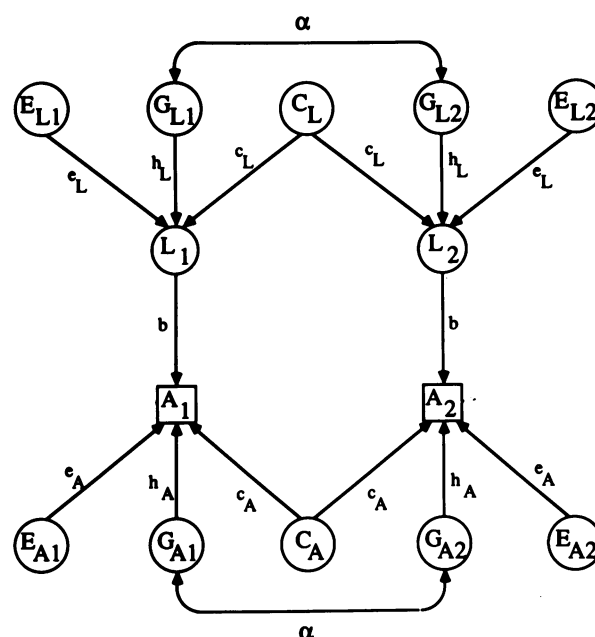


Figure 4 Path model showing resemblance for liability (L_i) and age at onset (A_i) as a function of additive genetic (G_{Li} and G_{Ai}), shared environmental (C_L and C_A), and specific environmental (E_{Li} and E_{Ai}) variation. The path coefficient α is not estimated; it is fixed at 1.0 for MZ twins and at 0.5 for DZ twins.

to be approximately 94% accurate when compared with blood typing. However, zygosity diagnosis remains unknown in 15.3% of twin pairs. Assessment of psychiatric illness was obtained from medical records of active military service, claims for Veteran's Administration (VA) disability, medical care at any VA facility, health questionnaires, and death certificates. Although the diagnoses were made by a variety of American clinicians from the 1940s to the 1970s, they were made in clinical situations unrelated to any research hypotheses. Unfortunately, the diagnosis cannot be regarded as accurate, and the possibility of misclassification in both directions exists. The possible effects of such misclassification on the parameter estimates will be discussed.

From records updated in 1981, 194 MZ and 277 DZ twin pairs in which at least one twin had a diagnosis of schizophrenia were found. With the exception of one twin pair, any recorded diagnosis of schizophrenia was accompanied by the date of first diagnosis, which is the age at onset used in the present analysis. The youngest twins in the registry would have been 54 years old in 1981, and therefore the population may be assumed

Table 3

Summary Statistics for Age at Onset in the NAS-NRC Schizophrenic Twin Sample

	MZ		DZ	
	Concordant	Discordant	Concordant	Discordant
N	30	134	9	259
Mean	29.72	34.42	34.06	32.97
SD	9.20	10.88	10.06	10.26

to have passed through the age of risk for schizophrenia, thus avoiding the additional complication of censoring of the data. However, owing to early death, approximately 5.5% of co-twins would not have fully completed the age of risk for schizophrenia.

Model Fitting

Summary statistics for the NAS-NRC schizophrenia data are shown in table 3. The superabundance of concordant MZ pairs (22.4%, vs. 3.5% for DZ pairs) suggests a substantial genetic component to familial resemblance for liability to disease. In the MZ twins, we see a lower mean age at onset in concordant pairs, suggesting a negative correlation between liability to disease and age at onset. This pattern is not replicated in the DZ twin pairs, but the sample size of concordant pairs is very small for this zygosity. The reduced variance of age at onset for concordant pairs is also consistent with a relationship between liability to disease and age at onset.

The model shown in figure 4 was fitted to the data

on affection status and age at onset. The population cumulative lifetime incidence was set at 1.8%, the observed frequency of schizophrenia in the NAS-NRC Twin Registry. The figure is high and reflects the relatively broad diagnostic criteria used in the present study. Parameter estimates and log likelihoods for both the full model and a number of submodels are shown in table 4.

Model I is not constrained and indicates a substantial additive genetic component to liability to schizophrenia, as has been reported elsewhere (Kendler 1983). There is also a moderate causal relationship between liability to disease and age at onset, as well as some additive genetic variation for specific age-at-onset factors which are independent of liability. Estimates of common environmental influences are at their lower bound of zero for both liability and the specific age-at-onset factors.

Twice the difference in log likelihoods between submodels is approximately distributed as χ^2 , with df equal to the number of parameters fixed in the submodel. When additive genetic effects on liability to disease are fixed at zero, a highly significant difference of 26.21 is obtained. However, when the causal relationship between age at onset and liability is removed from the model, no significant deterioration in fit is observed. While there is significant familial resemblance for age-at-onset factors (model IV), the hypothesis that this resemblance is due to shared environmental factors cannot be rejected (model V). By Akaike's (1987) information criterion the most parsimonious model (model III) is one with additive genetic

Table 4

Parameter Estimates of Genetic and Environmental Sources of Variation for Liability to and Age at Onset of Schizophrenia in the NAS/NRC Sample ($N_{MZ} = 164$; $N_{DZ} = 268$)

PARAMETER	MODEL				
	I	II	III	IV	V
h_L	.8383	.83	.83
c_L	.00	.71	.00	.00	.00
b	-.43	-.40	...	-.51	-.42
h_A	.69	.70	.76
c_A	.00	.00	.0058
μ	45.0	44.1	33.2	47.83	47.1
σ	11.3	11.2	10.4	11.8	11.7
-2 Log L	4,309.05	4,335.26	4,310.35	4,317.76	4,312.09
χ^2 Difference ^a	...	26.21	1.30	8.71	3.04

^a Difference in function value of models II-V versus model I.

and random environmental variation in both schizophrenia liability and age at onset, with no causal relationship between them.

It is of interest to consider the estimates of the mean of the age at onset of schizophrenia. When no causal relationship between age at onset and liability is specified, the estimate of mean age at onset is close to the observed mean for all affected individuals (33.01). However, when liability to disease causes variability in age at onset, the observed mean age at onset is no longer a good estimate of the true population mean. Truncation on the liability dimension results in truncation on the age-at-onset distribution when the two are correlated. Thus, for negative estimates of b , the estimate of the mean is expected to exceed the observed sample mean, and, for positive estimates of b , the estimated mean is expected to be less than the observed sample mean.

Discussion

The use of age at onset as an index for disease liability has been explored. The mathematical treatment is based on conventional likelihood methods for the analysis of indexes of liability to illness and is modified to allow for the lack of observation of age at onset in unaffected individuals. Results from simulation studies are encouraging but suggest that sample sizes need to be large, especially for less common phenotypes, in order to estimate parameters of the model precisely. Perhaps the most useful applications of these methods will be with regard to common disorders such as alcohol abuse, smoking, phobias, or depression. Age at onset is one example of an index of disease liability that can be measured only in affected individuals. The likelihood expressions are general for any such index. Clearly, the most informative indexes are those which may be assessed in both affected and normal individuals. Indexes that can be measured only in normal individuals will be more useful for disorders with a low cumulative lifetime incidence, since concordant affected pairs will be extremely rare.

Analysis of data collected under a number of ascertainment methods would seem impossible with current software and hardware. However, some improvement in numerical integration may be obtained by using faster, less general routines such as MULNOR (Schervish 1984) or the more recent QAVNOR (Baigorri et al., submitted). Considerable improvement could be obtained by using an approximation (e.g., see Rice et al. 1979), but the net effect of inaccuracies on the likelihood surface is difficult to assess. Superior computer

architecture with more-efficient parallel processing, such that nondependent subroutine calls within loops may be processed in parallel, would help to make integration over a larger number of dimensions feasible.

Application to Schizophrenia Data

A simple model of familial resemblance, allowing for additive genetic and for both common and specific environmental variation in liability to disease and age at onset, as well as for a causal relationship between liability to disease and age at onset, was fitted to the NAS-NRC twin data. Results indicate a large additive genetic component to liability to disease. Although the parameter estimate of the relationship between disease liability and age at onset is quite large, it is nonsignificant by likelihood ratio test. There is evidence for independent factors influencing age at onset that correlate between relatives. This latter finding is important for two reasons. First, there are implications for the prediction of risk to relatives according to the age at onset in the proband. Under the best-fitting model, these risks have a very similar pattern to those shown in the table (model 2). Second, there is the possibility that treatments that postpone onset of the disease may be devised. Postponement beyond the lifespan of the individual would effectively prevent the disease.

While the results from fitting the model to the NAS-NRC data are interesting, some caution with interpretation is required. The possibility of inconsistent diagnosis has a number of implications. If truncation of the phenotype is not occurring according to a fixed threshold but follows some probability distribution according to the position on the liability scale, then one of two forms of selection may be operating. First, there may be simply some error of measurement with respect to diagnosis, error that is not correlated between relatives. This is "indirect soft selection" of the form described by Martin and Wilson (1982). The consequences of such selection are to attenuate correlations further, and thus the estimates of familial resemblance for liability to disease and age at onset may be underestimated. Second, soft selection may be of the direct form described by Neale et al. (1989), in which the probability of being diagnosed varies directly as a function of the liability. Under this latter form, correlations between disease liabilities of relatives, and the correlations with other variables, are expected to be less attenuated than those given by truncate selection. Thus, parameters estimated under a truncate selection model may be overestimated; for schizophrenia, there may be no substantial familial aggregation for age-at-onset factors that are independent of disease liability. To distinguish be-

tween the two forms of selection, comparisons of familial correlations for indexes may be compared when the criteria for diagnosis are varied, but the power to discriminate may be low.

A second problem with the NAS-NRC data is that not all the twins in the sample have passed the age of risk. In 1981 the minimum age of living twins in the registry was 54 and was likely to be beyond age at onset for almost all twins, but some twins may have died before reaching age at onset (Kendler and Robinette 1983). If age at onset and liability are not independent, then we may expect that individuals lower on the liability scale are being excluded at a higher rate than are those with high liabilities. This selection would tend to bias both against the establishment of a relationship between liability and onset and against finding discordant pairs, inflating the correlation for liability to disease. These effects are expected to be small for the NAS-NRC data, as the age-corrected concordance calculated by Kendler and Robinette is only slightly different from the uncorrected concordance. However, the relationship between age at onset and disease liability may have been attenuated further if early-onset concordant pairs were not ascertained.

Estimating the Population Threshold

The treatment given in the present paper assumes that there is an estimate of the population threshold for disease that is error free. Unfortunately, such estimates are rarely if ever available in practice. Specification of a fixed threshold reduces the error on parameters of the model that covary with the threshold. Hence, tests of significance of these other parameters will be oversensitive. Two methods of estimating the population threshold are available, i.e., sampling at random and sampling groups of three or more relatives (Rice and Reich 1985). Random sampling may prove inefficient when the cumulative lifetime incidence is low.

The source of the information in larger groups of relatives is clear when we consider the pairs of relatives of a proband. These pairs may be concordant affected, discordant, or concordant normal, yielding information about population frequency as well as about covariance among relatives. Unfortunately, for each additional relative in the pedigree, it is necessary to add a further dimension of integration to compute the likelihood, which introduces considerable computing difficulties.

Multivariate Extensions of the Model

One of the most tractable extensions to the model is the addition of further indexes of liability. An in-

crease in the number of dimensions of the multinormal function does not result in great increases in computer time in this case, since integration is only needed for the liability dimensions. Some indexes of disease, such as age at onset or severity, may be measured only in affected individuals, while others, such as biochemical markers and physiological or psychological tests, may be measured in both normal and affected persons. Yet others may only reasonably be measured in normals, as medication or disease status may make testing inappropriate. While additional computer time is required to invert the covariance matrix and to estimate parameters of these distributions, these additions are relatively minor for potentially major improvements in the understanding of liability to disease.

Alternative Models of Age at Onset and Liability

The model presented and applied in the present paper is simple and testable, but it is not the only plausible model for onset and liability data. Within the general paradigm of linear models under the multifactorial assumption, several alternative relationships between disease liability and age at onset may be considered. First, age at onset could be specified to cause variation in disease liability. In principle, with cross-sectional data from twins reared together, this somewhat counterintuitive model may be discriminated from the liability-causes-age-at-onset model so long as the MZ and DZ twin correlations do not have the same values for the liability variable as they do for age at onset (Heath et al. 1989). The power to discriminate between these two models may be low in many cases. Second, both causal models may be compared against a general factor model. Here, each of the latent variables (additive genotype and both common and specific environment) could be partitioned into three components, i.e., specific to disease liability, specific to age at onset, and general (causing variation in both phenotypes). Third, we may consider submodels of the general factor model, for example by allowing all the covariation between age at onset and liability to be due to genetic factors alone.

For some variables, a shift of paradigm away from linear modeling may be appropriate. Models of survival analysis that have been extended to allow correlation between relatives for time to failure (onset), such as that described by Meyer and Eaves (1988), may be compared empirically for any set of data. Both approaches may be extended beyond a multifactorial treatment of variation to allow for the effects of major genes or for a combination of these sources under a mixed model.

Acknowledgments

This research was supported in part by NIH grants AG-04954 and GM-30250 and by ADAMHA grants MH-40828, MH-41953, GM-30250, and AA-06781. The NAS-NRC Twin Registry is maintained through contract HV-12919 with the National Heart, Lung and Blood Institute. The use of military, Veterans Administration, and NAS-NRC Twin Registry records in this research is acknowledged but is not to be construed as implying official approval of the conclusion by the departments and agencies that provided these records. The conclusions presented in this article do not necessarily represent those of the NAS-NRC Twin Registry. The authors are very grateful for helpful comments by A. C. Heath, M. L. Marazita, E. Simonoff, and two anonymous reviewers of the manuscript.

References

- Aitken AC (1934) Note on selection from a multivariate normal population. *Proc Edinb Math Soc [B]* 4:106–110
- Akaike H (1987) Factor analysis and AIC. *Psychometrika* 52:317–332
- Anderson DE (1972) A genetic study of human breast cancer. *JNCI* 48:1029–1034
- Baigorri AR, Van Eerdewegh P, Reich T. Error bounded integration of the multivariate normal densities over rectangular regions (submitted)
- Boehnke M, Lange K (1984) Ascertainment and goodness of fit of variance component models for pedigree data. In: Rao DC, Elston RC, Kuller LH, Feinleib M, Carter C, Havlik R (eds) *Genetic epidemiology of coronary heart disease: past, present, future*. Alan R. Liss, New York, pp 173–192
- Bucher KD, Schrott HG (1982) The effect of nonrandom sampling on familial correlations. *Biometrics* 38:249–253
- Cloninger CR, Christiansen KO, Reich T, Gottesman H (1978) Implications of sex differences in the prevalences of antisocial personality, alcoholism, and criminality for familial transmission. *Arch Gen Psychiatry* 35:941–951
- Curnow RN (1972) The multifactorial model for the inheritance of liability to disease and its implications for relatives at risk. *Biometrics* 28:931–946
- (1974) The use of additional information in estimating disease risk from family histories. *Biometrics* 30: 655–665
- Curnow RN, Smith C (1975) Multifactorial models for familial diseases in man. *J R Stat Soc [A]* 13:131–169
- Eaves LJ (1970) Aspects of human psychogenetics. PhD thesis, University of Birmingham
- (1983) Errors of influence in the detection of major gene effects on psychological test scores. *Am J Hum Genet* 35:1179–1189
- Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. *Hum Hered* 21:523–542
- Falconer DS (1965) The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann Hum Genet* 29:51–76
- Fisher RA (1922) On the mathematical foundations of theoretical statistics. *Philos Trans R Soc Lond [A]* 22:309–368
- (1931) *British Association's mathematical tables*. British Association for the Advancement of Science, Cambridge, vol 1, pp xxvi–xxxv
- Heath AC, Berg K, Eaves LJ, Solaas MH, Corey LA, Sundet J, Magnus P, et al (1985) Education policy and the heritability of educational attainment. *Nature* 31:734–736
- Heath AC, Jardine R, Eaves LJ, Martin NG. The genetic structure of personality. II. Genetic item analysis of the EPQ. *J Pers Individ Differences* (in press)
- Heath AC, Neale MC, Hewitt JK, Eaves LJ, Fuller DW (1989) Testing structural equation models for twin data using LISREL. *Behav Genet* 19:9–35
- Heiberg A, Slack J (1977) Family similarities in the age at coronary death in familial hypercholesterolaemia. *Br Med J* 2:493–495
- Heimbuch RC, Matthyse S, Kidd KK (1980) Estimating age-of-onset distributions for disorders with variable onset. *Am J Hum Genet* 32:564–574
- Hewitt JK, Neale MC. A simple simulation procedure for exploring genetic models. *Behav Genet* (in press)
- Jablon S, Neel JV, Gershowitz H, Atkinson GF (1967) The NAS-NRC twin panel: methods of construction of the panel, zygosity diagnosis, and proposed use. *Am J Hum Genet* 19:133–161
- Jinks JL, Fulker DW (1970) Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychol Bull* 73:311–349
- Johnson NL, Kotz S (1972) *Distributions in statistics: continuous multivariate distributions*. Wiley, New York
- Kendler KS (1983) Overview: a current perspective on twin studies of schizophrenia. *Am J Psychiatry* 140:1413–1425
- Kendler KS, Robinette CD (1983) Schizophrenia in the National Academy of Sciences–National Research Council Twin Registry: a 16-year update. *Am J Psychiatry* 140:1551–1563
- Kendler KS, Tsuang MT, Hays P (1987) Age at onset in schizophrenia: a familial perspective. *Arch Gen Psychiatry* 44:881–890
- Larsson T, Sjögren T (1954) A methodological, psychiatric and statistical study of a large Swedish rural population. *Acta Psychiatr Neurol Scand (Suppl)* 89:40–54
- McGue M, Gottesman II, Rao DC (1983) The transmission of schizophrenia under a multifactorial threshold model. *Am J Hum Genet* 35:1161–1178
- Martin NG, Eaves LJ, Kearsley MJ, Davies P (1978) The power of the classical twin study. *Heredity (Edinburgh)* 40:97–116
- Martin NG, Wilson SR (1982) Bias in the estimation of heritability from truncated samples of twins. *Behav Genet* 12: 467–472
- Meyer JS, Eaves LJ (1988) Estimating genetic parameters of

- survival distributions: a multifactorial model. *Genet Epidemiol* 5:265–275
- Neale MC (1988) Handedness in a sample of volunteer twins. *Behav Genet* 18:69–79
- Neale MC, Eaves LJ, Kendler KS, Hewitt JK (1989) Bias in correlations from selected samples of relatives: the effects of soft selection. *Behav Genet* 19:163–169
- Neale MC, Rushton JP, Fulker DW (1986) The heritability of items from the Eysenck personality questionnaire. *Pers Individ Differences* 7:771–779
- Numerical Algorithms Group (1988) Numerical Algorithms Group FORTRAN library manual, mark 12. Numerical Algorithms Group, Oxford
- Pearson K (1900) Mathematical contributions to the theory of evolution. VIII. On the correlation of characters not quantitatively measurable. *Proc R Soc Lond* 66:316–323
- Pericak-Vance MA, Elston RC, Conneally PM, Dawson DV (1983) Age-of-onset heterogeneity in Huntington disease families. *Am J Med Genet* 14:49–59
- Price RA, Kidd KK, Weissman MM (1987) Early onset (under age 30 years) and panic disorder as markers for etiologic homogeneity in major depression. *Arch Gen Psychiatry* 44:434–440
- Rao DC, Wette R (1987) Nonrandom sampling in genetic epidemiology: maximum likelihood methods for multifactorial analysis of quantitative data ascertained through truncation. *Genet Epidemiol* 4:357–376
- Rao DC, Wette R, Ewens WJ (1988) Multifactorial analysis of family data ascertained through truncation: a comparative evaluation of two methods of statistical inference. *Am J Hum Genet* 42:506–515
- Rice J, Reich T (1985) Familial analysis of qualitative traits under multifactorial inheritance. *Genet Epidemiol* 2:301–315
- Rice J, Reich T, Cloninger CR, Wette R (1979) An approximation to the multivariate normal integral: its application to multifactorial qualitative traits. *Biometrics* 35:451–459
- Ridley RM, Baker HF, Crow TJ (1986) Transmissible and non-transmissible neurodegenerative disease: similarities in age of onset and genetics in relation to aetiology. *Psychol Med* 16:199–207
- Risch N (1983) Estimating morbidity risks with variable age of onset: review of methods and a maximum likelihood approach. *Biometrics* 39:929–939
- Rissanen AM (1979) Familial occurrence of coronary heart disease: effect of age at diagnosis. *Am J Cardiol* 44:60–66
- Schervish MJ (1984) Multivariate normal probability with error bounded. *Appl Stat* 33:81–94
- Smith C, Mendell NR (1974) Recurrence risks from family history and metric traits. *Ann Hum Genet* 37:275–286
- Strömberg E (1935) Zum Ersatz des Weinbergischen "abgekürzten Verfahrens": Zugleich ein Beitrag zur Frage von der Erbllichkeit des Erkrankungsalters bei der Schizophrenie. *Z Gesamte Neurol Psychiatr* 153:784–797
- (1938) Beiträge zur psychiatrischen Erblehre auf Grund von Untersuchungen an einer Inselbevölkerung. *Acta Psychiatr Neurol Scand (Suppl)* 19:1–257
- Sturt E (1986) Application of survival analysis to the inception of dementia. *Psychol Med* 16:583–593
- Weinberg W (1925) Methoden und Technik der Statistik mit besonderer Berücksichtigung der Socialbiologie. In: Gottstein A, Schlossman A, Teleky L (eds) *Handbuch der sozialen Hygiene und Gesundheitsfürsorge*. Springer, Berlin, pp 71–148
- Wright S (1934) The method of path coefficients. *Ann Math Stat* 5:161–215