Acta Genet Med Gemellol 29:137–142 (1980)
The Mendel Institute/Alan R. Liss, Inc.

Received 6 November 1979
Final 31 January 1980

Conditional Concordance in Monozygotic Twins

Brian Suarez

Department of Psychiatry, Washington University School of Medicine; The Jewish Hospital of St. Louis, Missouri

Twin concordance rates are usually reported without reference to the number of parents affected, apparently because the simple demonstration that monozygotic (MZ) twins are more concordant than are dizygotic (DZ) twins is the goal of most twin studies. Depending on the underlying mechanism, however, twin concordance rates can vary widely when conditioned on the number of parents affected. For the generalized single-locus model it is shown that conditional concordance rates in monozygotic twins, along with an estimate of the disorder's prevalence in the population, uniquely specify the underlying parameters of this important model. Knowledge of the exact parameter set is essential for competent genetic counseling.

Key words: Twin concordance rates, Conditional concordance, Single-locus model, Genetic counseling

INTRODUCTION

One of the most persuasive pieces of evidence implicating transmissible factors in the etiology of physical or behavioral disorders is the demonstration that monozygotic (MZ) twins are more concordant than dizygotic (DZ) twins. Indeed, this type of evidence is so persuasive that in many cases it seems to be the goal of twin research rather than the starting point for further investigation. Increased concordance rates for MZ twins compared to DZ twins, by themselves, do little to elucidate the exact mode of transmission and, surely, do little to address what Elston and Stewart [1] assert to be the single most important question concerning the inheritance of any character: " . . . is most of the genetic variation due to one locus or are many loci necessarily involved?"

When a disorder is suspected of resulting from allelic variation at an incompletely penetrant single locus, however, the reporting of unconditional twin concordance rates does not provide sufficient information to estimate the underlying parameters of this general model. In this paper it is shown that this ambiguity can be removed by conditioning concordance rates on the number of affected parents. When this approach is taken, conditional concordance rates from just MZ twins are sufficient to obtain unique estimates of the model's parameters.

0001-5660/80/2902-0137$01.40

THE MODEL

The generalized single-locus model assumes that the disorder under study is determined by allelic variation at a single autosomal locus. A description of the main features of the model can be found in Reich et al [6]. In the treatment given here, however, the assumption that the liability for the trait is normally distributed within a genotype can be dropped, since only one threshold is assumed. Although there may be multiple alleles in the population, only two qualitative classes are assumed, that is, for a system of alleles $A_1, A_2, \ldots, A_K$, all relevant variation must be describable as A_1, say, and non-A_1. Since all non-A_1 alleles are assumed to be equivalent, in effect we denote the entire collection by A_2. Let p denote the population frequency of A_1 and $q = 1 - p$ the frequency of all remaining alleles. Now suppose that a proportion f_i of each genotype is affected with the disorder, where $i = 1, 2, 3$, for the respective genotypes A_1A_1, A_1A_2, A_2A_2. With panmixia, the prevalence of the disorder in the general population (K) is given by:

$$K = p^2 f_1 + 2pq f_2 + q^2 f_3 \tag{1}$$

The f_i are allowed to assume any values between zero and one and need not be ordered. Mendelian transmission is then a special case of this general model when $f_1 = 0$, $f_3 = 1$, and $f_2 = 1$ or 0, depending on whether the disorder is dominant or recessive.

The model's four parameters (f_1, f_2, f_3, q) can be used to define two useful quantities, the additive variance of each locus (V_A) and its dominance variance (V_D) [5]. These are, respectively:

$$V_A = 2pq\,[q(f_3 - f_2) + p(f_2 - f_1)]^2 \tag{2}$$

$$V_D = p^2 q^2 (f_1 - 2f_2 + f_3)^2 \tag{3}$$

These variance components have the same meaning for the generalized single-locus model that they have for polygenic traits. That is, V_A measures the variance of breeding values and is the chief determinant of how much a population will change if it is subjected to directional selection, whereas V_D measures the variance contributed when the proportion of affected heterozygotes is not exactly halfway between the proportions of the two homozygotes.

The three quantities, K, V_A, and V_D are important parameters because, among other things, they determine the unconditional concordance expected for any class of relatives of a proband. For MZ co-twins, the concordance rate (C) is simple [4]:

$$C = K + \frac{V_A + V_D}{K} \tag{4}$$

The probandwise concordance rate for any class of relatives other than MZ twins is easily calculated from Eq. 4 by weighting V_A by the probability that persons in that relationship share a given allele identical by descent and by weighting V_D by the probability that they share both alleles identical by descent.

The three parameters that determine the probandwise concordance rate (ie: K, V_A, V_D) map into a continuum of parameter sets in the four-dimensional parameter space (ie: f_1, f_2, f_3, q) that defines the underlying generalized single-locus model. This indeterminism has been referred to as the "parameter problem" [2].

For many purposes, a researcher is content to obtain estimates of K, V_A, and V_D, since these allow an initial fit of the model to incidence data [8], or, alternatively, can be used in a linkage analysis [7, 10]. Once the model's adequacy has been established, however, it is desirable to obtain an estimate of the four underlying parameters, especially if the analysis is to be used for counseling purposes. Two different approaches for obtaining unique estimates of q, f_1, f_2, and f_3 have been suggested. When two separate populations are known to differ in K, then the methods of Suarez et al [9] may be used. Alternatively, unique estimates may be obtained from a single population by conditioning the probandwise concordance rate on an additional observation [2]. Any class of relatives may be used to obtain these estimates, although the further removed they are from the proband, the poorer the estimates. Since MZ twins are, by definition, genetically identical, they are capable of providing the most information. For MZ twins the most convenient additional conditioning event is the number of parents similarly affected with the disorder. This information is often available to a researcher, but rarely is it reported.

Derivation of the conditional concordance rate for MZ twins can be facilitated by considering the following three-by-three array:

Number of parents affected	*Proband's genotype*			
	A_1A_1	A_1A_2	A_2A_2	
Both	ω_{21}	ω_{22}	ω_{23}	$\omega_{2.} = K^2K_2$
One	ω_{11}	ω_{12}	ω_{13}	$\omega_{1.} = 2K(1-K)K_1$
Neither	ω_{01}	ω_{02}	ω_{03}	$\omega_{0.} = (1-K)^2 K_0$
	$\omega_{.1} = p^2 f_1$	$\omega_{.2} = 2pqf_2$	$\omega_{.3} = q^2 f_3$	$\omega_{..} = K$

The entry in each cell (the ω_{ij}) is the probability that a person has genotype j(j = 1, 2, 3, for the respective genotypes A_1A_1, A_1A_2, A_2A_2), is affected, and has i affected parents (where i = 0, 1, or 2, depending on whether neither, one, or both parents are affected, respectively). Except for mendelian transmission, these quantities are not directly observable since, with incomplete penetrance, there is no one-to-one transformation between genotype and phenotype. The ω_{ij} can be thought of as the distribution of cells from which the probands are drawn. Because the genotype of a proband is not directly observable, the marginal totals, $\omega_{.j}$, are not either. The marginal totals, $\omega_{i.}$, are the probabilities of observing an affected child (the proband) and i affected parents. These quantities are directly observable, being simply the proportion of affected offspring given i affected parents weighted by the probability of i affected parents. Assuming random mating, the probability of 2, 1, or 0 affected parents is K^2, $2K(1-K)$, and $(1-K)^2$, respectively. The proportion of affected children, given i affected parents (the K_i), is given in Suarez et al [9]. Since these proportions can be expressed in terms of just K, V_A, V_D, and since the probability of observing i affected parents depends only on K, the $\omega_{i.}$ are not dependent on the four unique parameters of the model (ie, q and f_i). Associated with each $\omega_{i.}$ is the conditional concordance rate, C_i, directly observable, although rarely reported. These are, of course, obtained by determining whether a proband's co-twin is affected, and since the co-twin is identical to the proband, for any row in the array the expected concordance rate is simply:

$$(f_1\omega_{i1} + f_2\omega_{i2} + f_3\omega_{i3})/\omega_{i.}$$

https://doi.org/10.1017/S0001566000008618 Published online by Cambridge University Press

Consequently, the C_i depend on all four underlying parameters as given below:

$$C_2 = [p^2 f_1^2 (p^2 f_1^2 + 2pq\, f_1 f_2 + q^2 f_2^2) + 2pqf_2^2 (p^2 f_1 f_2 + pqf_1 f_3 + pqf_2^2 + q^2 f_2 f_3) + q^2 f_3^2 (p^2 f_2^2 + 2pqf_2 f_3 + q^2 f_3^2)] / \omega_{2.}, \quad (5)$$

$$C_1 = \langle p^2 f_1^2 [2p^2 f_1 (1 - f_1) + 2pq\, [f_1 (1 - f_2) + f_2 (1 - f_1)] + 2q^2 f_2 (1 - f_2)] + 2pq f_2^2 [p^2 [f_1 (1 - f_2) + f_2 (1 - f_1)] + pq[f_1 (1 - f_3) + f_3 (1 - f_1)] + 2pqf_2 (1 - f_2) + q^2 [f_2 (1 - f_3) + f_3 (1 - f_2)]] + q^2 f_3^2 [2p^2 f_2 (1 - f_2) + 2pq[f_2 (1 - f_3) + f_3 (1 - f_2)] + 2q^2 f_3 (1 - f_3)] \rangle / \omega_{1.}, \quad (6)$$

$$C_0 = \langle p^2 f_1^2 [p^2 (1 - f_1)^2 + 2pq(1 - f_1)(1 - f_2) + q^2 (1 - f_2)^2] + 2pqf_2^2 [p^2 (1 - f_1)(1 - f_2) + pq(1 - f_1)(1 - f_3) + pq(1 - f_2)^2 + q^2 (1 - f_2)(1 - f_3)] + q^2 f_3^2 [p^2 (1 - f_2)^2 + 2pq(1 - f_2)(1 - f_3) + q^2 (1 - f_3)^2] \rangle / \omega_{0.}. \quad (7)$$

Let the observed probandwise concordance rate be denoted as 0_2, 0_1, and 0_0 for MZ twins with both, one, and neither parent affected, respectively. The parameters q and f_i are then obtained by maximizing the function:

$$F = -\Sigma((0_2 - C_2)^2 + (0_1 - C_1)^2 + (0_0 - C_0)^2 + (\hat{K} - K)^2)$$

where $\hat{K}$ is the predicted population prevalence from Eq. 1.

The surface described by this function in terms of the parameters q and f_i, appears to contain many local maxima, making the search procedure tedious and convergence slow. However, when the search is given new parameters by substituting the following three equations for the f_i, the surface appears smooth, and convergence is rapid. The substitutions are [8]:

$$f_1 = K - \frac{\sigma_A (2pq)^{1/2} + \sigma_D q}{p} \quad (8)$$

$$f_2 = K - \sigma_D + \frac{\sigma_A (1 - 2q)}{(2pq)^{1/2}} \quad (9)$$

$$f_3 = K + \frac{\sigma_A (2pq)^{1/2} + \sigma_D p}{q} \quad (10)$$

where $\sigma_A = (V_A)^{1/2}$ and $\sigma_D = \pm(V_D)^{1/2}$.

DISCUSSION

It is important to reemphasize that, under the assumptions of the generalized single-locus model treated here, a person's probability of being affected depends only on that person's genotype. In other words, expression of the trait is not influenced by the particular genetic background or environmental circumstances present. This problem of incomplete penetrance has concerned researchers for many years. An early attempt at estimating penetrance from twin concordance rates was made by Huizinga and Heiden [3]. The model they investigated is a special case of the one treated here. In the terminology of the present paper, Huizinga

and Heiden assume that $f_1 = 0$ and $f_3 = 1$. They then estimate the values of f_2 and the gene frequency by comparing the unconditional concordance rate in DZ twins with that of MZ twins. In contrast, the approach taken here is to estimate all four parameters by fitting a less restrictive model to the conditional concordance rates of MZ twins.

It is indeed unfortunate that so few workers report the conditional concordance rates for MZ or, for that matter, DZ twins. The reason for this failure most likely lies in the fact that the goal of much twin research is the simple demonstration of transmissible factors, which is easily accomplished by comparing the unconditional concordance rate of MZ and DZ twins. As a result, conditional concordance rates have largely been ignored. While it is the case that the frequency of MZ twins with one or both parents affected may represent a small proportion of a total twin sample, this need not deter their being reported.

As an example of the power of conditional concordance rates in MZ twins to resolve the "parameter problem," consider a disorder with a population prevalence of K = 5%. Suppose further that the additive and dominance variances are both 0.0064. These parameters yield an unconditional concordance rate for MZ twins of 30.6% which, when compared to the unconditional concordance rate of 14.6% for DZ twins, argues for a moderate heritable component (h^2 in the broad sense is 26.95% for this example). These unconditional concordance rates are consistent with a broad spectrum of parameters in the $\langle q, f_i \rangle$ space. With techniques developed elsewhere [9], the lower limit of this spectrum is found to occur at a gene frequency of $q = 0.112$, at which point the penetrances are $f_1 = 0.0199$, $f_2 = 0.1089$, $f_3 = 1.0$. The upper limit occurs at a gene frequency of $q = 0.372$, where $f_1 = 0.0103$, $f_2 = 0$, $f_3 = 0.3322$. Both of these parameter sets and all intermediate sets defined by Eq. 8–10 with $0.112 < q < 0.372$ will yield the same unconditional concordance rates in MZ twins. However, as the Figure indicates, the conditional concordance rates differ widely, depending on the exact set of underlying parameters. For this example, the conditional concordance rates are most different when q is small (0.112) and, for practical purposes, indistinguishable as q approaches its upper limit (0.372). Note that the frequency with which neither parent, one parent, or both parents are affected is independent of the exact $\langle q, f_i \rangle$ parameter set. For the above example, only 1.4% of parental couples are both expected to be affected. However, 20% of all twin pairs will have one parent who is affected.

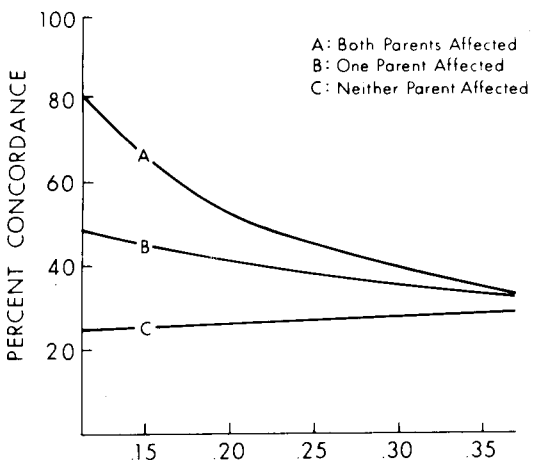

Figure. Probandwise concordance rate in MZ twins with (A) both parents, (B) one parent, (C) neither parent affected. The unconditional concordance is 30.6% for all gene frequencies (and appropriate penetrances) in the range $0.112 < q < 0.372$ since all are capable of yielding $K = 0.05$ and $V_A = V_D = 0.0064$.

https://doi.org/10.1017/S0001566000008618 Published online by Cambridge University Press

Knowledge of the exact gene frequency and penetrance set is essential for adequate genetic counseling. In the present example, for instance, the recurrence risk for a family that has already segregated an affected child is 54.3% when both parents are affected, and the exact parameter set is at its lower bound (ie, when q = 0.112); whereas for the same family the recurrence risk is 33.2% if the parameter set is at its upper bound (ie, q = 0.372). Since unusual families are most likely to seek genetic counseling, the potential leverage gained from MZ twin conditional rates in determining the exact parameter set will be welcome.

Acknowledgments. It is my pleasure to express my gratitude to Dr. John Rice for his help in reparameterizing the computer search, and to Drs. T. Reich, I. Gottesman, and G. Carey for their helpful comments, to Nancy Caston for her excellent programming assistance, and to Ms Lora Loggins for her help in preparation of the manuscript. This work has been supported, in part, by USPHS grants MH-31302 and MH-14677.

REFERENCES

1. Elston RC, Stewart J (1971): A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542.
2. Fishman PM, Reich T, Suarez B, James JW (1978): A note on the essential parameters of the two-allele autosomal locus model. Am J Hum Genet 30:283–292.
3. Huizinga J, Heiden JA (1957): The percentages of concordance in twins and mode of inheritance. Acta Genet Med Gemellol 6:437–450.
4. James JW (1971): Frequency in relatives for an all-or-none trait. Ann Hum Genet 35:47–49.
5. Kempthorne O (1957): "An Introduction to Genetic Statistics." New York: John Wiley.
6. Reich T, James JW, Morris CA (1972): The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. Ann Hum Genet 36:163–184.
7. Suarez B (1978): The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. Tissue Antigens 12:87–93.
8. Suarez BK, Reich T, Trost J (1976): Limits of the general two-allele single locus model with incomplete penetrance. Ann Hum Genet 40:231–244.
9. Suarez B, Fishman PM, Reich T (1977): Estimating the parameters of the incompletely penetrant single locus model using multiple populations. Hum Hered 27:336–351.
10. Suarez B, Rice J, Reich T (1978): The generalized sib pair IBD distribution: Its use in the detection of linkage. Ann Hum Genet 42:87–94.

Correspondence: Dr. B. K. Suarez, Department of Psychiatry, The Jewish Hospital, PO Box 14109, St. Louis, MO 63178, USA.