

## Twin Concordance Rates

Two different concordance rates are often used in twin studies in which twins are not sampled from the general population, the pairwise rate and the probandwise rate. Actually, these two concordance rates are special cases of a general, maximum likelihood method for calculating concordance that can easily be done on any twin data. The only key to using the general formula is to keep track of the doubly ascertained twin pairs during the data collection phase of a study. First, I present the algebraic formulae, then a numerical example, and finally provide the proof.

### The algebraic formulae

Let  $C_1$  denote the number of concordant pairs in the sample where both twins were independently ascertained. Note that  $C_1$  is the number of *pairs* and not the number of *probands*. For example, if both members of the Smith twins and both members of the Jones twins are doubly ascertained, then  $C_1 = 2$  and not 4. Let  $C_2$  denote the number of concordant pairs in which one and only one member was ascertained and let  $C = C_1 + C_2$  or the total number of concordant pairs. Finally, let  $D$  denote the number of discordant pairs.

The pairwise concordance rate, denoted here as  $A$ , is

$$A = \frac{C}{C + D}. \quad (1)$$

The proband concordance rate, denoted as  $B$  is

$$B = \frac{2C_1 + C_2}{2C_1 + C_2 + D}. \quad (2)$$

The first step in the approach outlined here is to calculate the quantity  $\pi$ , the probability of ascertaining an individual with the trait under study. This quantity can be estimated directly from the observed twin data by the equation

$$\hat{\pi} = \frac{2C_1}{2C_1 + C_2}.$$

Although one could calculate two estimates of  $\pi$ , one for MZ and the other for DZ twins, it is preferable to pool the estimate. Hence, the recommended equation becomes

$$\hat{\pi} = \frac{2(C_{1mz} + C_{1dz})}{2(C_{1mz} + C_{1dz}) + C_{2mz} + C_{2dz}} \quad (3)$$

where subscripts *mz* and *dz* denote zygosity.

Let  $C_i$  denote the total number of concordant pairs for the *i*th zygosity. That is  $C_i = C_{1i} + C_{2i}$ . With  $\pi$  estimated from equation (1), an unbiased estimate of the population concordance rate ( $\theta_i$ ) is given by

$$\hat{\theta}_i = \frac{2C_i}{2C_i + (2 - \pi)D_i} \quad (4)$$

Naturally there will be a separate estimate of  $\theta$  for MZ and DZ twins. The quantity  $\theta$  is a conditional probability that estimates the population probability that a cotwin will be affected given that his/her partner is affected. It is also equal to the segregation ratio. Hence, it has a much more important meaning in quantitative genetics than either the pairwise or probandwise concordance.

When the ascertainment probability is very low,  $\pi$  approaches 0 and equation (4) reduces to the pairwise rate given in equation (1). In complete ascertainment where all affected individuals are probands, then  $\pi = 1.0$ ,  $C_2 = 0$ , and equation (4) reduces to the probandwise rate in equation (2).

Curiously, testing the significance of the difference between MZ and DZ concordance rates uses only the pairwise rates. First compute  $A_{mz}$  and  $A_{dz}$ —the pairwise concordance rates for MZ and DZ twins. Next, compute  $A$ , the pairwise concordance for all twins, ignoring zygosity:

$$A = \frac{C_{mz} + C_{dz}}{C_{mz} + C_{dz} + D_{mz} + D_{dz}}.$$

Finally, compute the likelihood ratio  $\chi^2$ :

$$\chi^2 = 2[C_{mz} \log(A_{mz}) + D_{mz} \log(1 - A_{mz}) + C_{dz} \log(A_{dz}) + D_{dz} \log(1 - A_{dz}) - C \log(A) - D \log(1 - A)] \quad (5)$$

This will be a  $\chi^2$  with one degree of freedom. Note that the logarithm taken here is the natural or Naperian logarithm.

### A numerical example

Gottesman & Shield's (1972) twin data on schizophrenia will be used to illustrate the procedure. The data used here consist of a consensus diagnosis of schizophrenia or probable/questionable schizophrenia derived from the independent evaluation of case histories by six clinicians (see Gottesman & Shields, 1972, Appendix C) and are given in Table 1.

[Insert Table 1 about here]

Using both MZ and DZ twins, the estimate of the ascertainment probability,  $\pi$ , is given from equation (3) as

$$\pi = \frac{2(4+1)}{2(4+1) + 7 + 2} = .526.$$

The concordance rate for identical twins is found by entering the quantities in Table 1 into equation (4), or

$$\theta_{mz} = \frac{2(11)}{2(11) + (2 - .526)11} = .576.$$

Similarly, the concordance for DZ twins is

$$\theta_{dz} = \frac{2(3)}{2(3) + (2 - .526)30} = .119.$$

To test for a significant difference in concordance, we compute the pairwise rates for the MZ twins, the DZ twins, and all twins regardless of zygosity. Thus  $A_{mz} = 11/22 = .50$ ,  $A_{dz} = 3/33 = .091$ , and  $A = 14/55 = .255$ . Using these quantities and equation (5), the likelihood ratio  $\chi^2$  is

$$\chi^2 = 2[11\log(.5) + 11\log(.5) + 3\log(.091) + 30\log(.909) - 14\log(.255) - 41\log(.745)] = 11.80$$

The value of  $\chi^2$  exceeds the critical value of 6.64 at the .01 level. Hence, there is clear evidence that concordance for schizophrenia is significantly greater in MZ than in DZ twins.

### Proof

Let  $p$  denote the prevalence of a trait in the general population. Under the assumption that the trait has the same prevalence in twins as in the general population the distribution of twins in the general population will be given in the simple two by two contingency table illustrated in Table 2.

[Insert Table 2 here]

Note that when twin pairs are randomly sampled from the general population, regardless of their trait status, then an estimate of  $\theta$  may be derived by simply double-entering the pairs into the table and dividing the proportion of concordant pairs by  $p$ .

When twins are ascertained through pairs where at least one twin has the trait, a correction for ascertainment is required to arrive at an unbiased estimate of  $\theta$  and it is necessary to estimate  $\pi$ . A critical assumption of the following derivation is that ascertainment is independent in members of concordant pairs. That is, it is assumed that the probability of independently ascertaining twin 2 in a concordant pair, given that twin 1 has already been ascertained, is  $\pi$ .

The probability of ascertaining twinships where neither partner has the trait is obviously 0. For discordant pairs, the probability of ascertaining the pair is simply  $\pi$ . For

concordant pairs, however, the ascertainment probability is more complicated. There are three ways to ascertain concordant pairs. First, twin 1 may be ascertained but twin 2 is not ascertained. This will occur with frequency  $\pi(1 - \pi)$ . Second, twin two may be ascertained while twin 1 remains unascertained. This also occurs with frequency  $\pi(1 - \pi)$ . Finally, both twins may be independently ascertained. This should occur with frequency  $\pi^2$ .

Using these quantities, one can then construct a table of the frequency with which twins should occur in the general population and in an ascertained sample. This table is given below in Table 3.

[Insert Table 3 here]

The quantity  $\lambda$  in the last column of table is the proportion of twin pairs in the general population that is ascertained. It may be found by summing the quantities in the sixth column (Frequency in the General Population by Ascertainment Status) of the table for all those pairs that are ascertained, or,

$$\lambda = \theta p \pi^2 + 2\theta p \pi(1 - \pi) + 2(1 - \theta) p \pi = (2 - \theta \pi) p \pi$$

Inspection of the last column in the table reveals that there are only three categories of twins in the ascertained twin sample. The first of these consists of concordant pairs in which both members are independently ascertained. Let  $C_1$  denote the number of these pairs in the sample and note that the probability of observing pairs of this type of twin pair in the sample is  $\theta p \pi^2 \lambda^{-1}$ . The second type are concordant pairs in which only one twin is ascertained. Let  $C_2$  denote these pairs. Their probability is  $2\theta p \pi(1 - \pi) \lambda^{-1}$ . The final type are discordant pairs. Their probability is  $2(1 - \theta) p \pi \lambda^{-1}$ , and their number will be denoted by  $D$ .

Consequently, the log likelihood of the sample becomes

$$\log(L) = C_1 \log(\theta p \pi^2 \lambda^{-1}) + C_2 \log[2\theta p \pi(1 - \pi) \lambda^{-1}] + D \log[2(1 - \theta) p \pi \lambda^{-1}]$$

which reduces to

$$\log(L) = cons + C_1 \log(\theta) + C_1 \log(\pi) + C_2 \log(\theta) + C_2 \log(1 - \pi) + D \log(1 - \theta) - N \log(2 - \theta\pi) \quad (6)$$

where *cons* is a constant and *N* is the total number of twin pairs ( $C_1 + C_2 + D$ ). Thus, the log likelihood is a function of two parameters,  $\theta$  and  $\pi$ . Differentiating the log likelihood with respect to each of these parameters and performing tedious algebra gives the maximum likelihood estimates of  $\pi$  and  $\theta$  as

$$\hat{\pi} = \frac{2C_1}{2C_1 + C_2}$$

and

$$\hat{\theta} = \frac{2(C_1 + C_2)}{2(C_1 + C_2) + (2 - \pi)D} = \frac{2C}{2C + (2 - \pi)D}.$$

Note that these are equations (3) and (4) given in the first section of this paper.

The likelihood ratio  $\chi^2$  is twice the difference in log likelihoods between a general model that fits three parameters to the data ( $\pi$ ,  $\theta_{mz}$ , and  $\theta_{dz}$ ) and a constrained model that fits  $\pi$  but constrains  $\theta_{mz}$  to equal  $\theta_{dz}$ . By subscripting the quantities in equation (6) to reflect MZ twins and DZ twins and then algebraically reducing the formula for  $\chi^2$ , one arrives at equation (5) given earlier. It is remarkable that all terms involving  $\pi$  cancel in the process and one is left with comparisons of the MZ, DZ, and whole-sample concordance rates.

## **References**

Gottesman, I.I. & Shields, J. (1972). *Schizophrenia and Genetics: A twin study vantage point*. New York: Academic Press.

Type:	Zygoty		
	MZ	DZ	All
$C_1$	4	1	5
$C_2$	7	2	9
$D$	11	30	41

Source: Gottesman & Shields (1972).

Table 2. Frequency of twins with (+) and without (-) a trait in the general population			
	Twin 2 +	Twin 2 -	Total
Twin 1 +:	$\theta p$	$(1 - \theta)p$	$p$
Twin 1 -:	$(1 - \theta)p$	$q - (1 - \theta)p$	$q$
Total	$p$	$Q$	1.0

Table 3. Distribution of Twin Types in the General Population and in an Ascertained Sample						
Twin Type						
Twin 1	Twin 2	Frequency in the General Population	Ascertainment Status	Probability of Ascertainment given Twin Type	Frequency in the General Population by Ascertainment Status	Frequency in the Ascertained Sample
+	+	$\theta p$	both ascertained	$\pi^2$	$\theta p \pi^2$	$\theta p \pi^2 \lambda^{-1}$
			twin 1 ascertained	$\pi(1 - \pi)$	$\theta p \pi(1 - \pi)$	$\theta p \pi(1 - \pi) \lambda^{-1}$
			twin 2 ascertained	$\pi(1 - \pi)$	$\theta p \pi(1 - \pi)$	$\theta p \pi(1 - \pi) \lambda^{-1}$
			neither	$1 - 2\pi(1 - \pi) - \pi^2$	$\theta p [1 - 2\pi(1 - \pi) - \pi^2]$	0
+	-	$(1 - \theta)p$	ascertained	$\pi$	$(1 - \theta)p \pi$	$(1 - \theta)p \pi \lambda^{-1}$
			not ascertained	$1 - \pi$	$(1 - \theta)p(1 - \pi)$	0
-	+	$(1 - \theta)p$	ascertained	$\pi$	$(1 - \theta)p \pi$	$(1 - \theta)p \pi \lambda^{-1}$
			not ascertained	$1 - \pi$	$(1 - \theta)p(1 - \pi)$	0
-	-	$q - (1 - \theta)p$	ascertained	0	0	0
			not ascertained	1	$q - (1 - \theta)p$	0