Hum. Hered. *32:* 202–207 (1982)

Ascertainment Bias for Non-Twin Relatives in Twin Proband Studies[1]

John Rice[a, b], *Irving I. Gottesman*[a, c], *Brian K. Suarez*[a, c], *Dennis H. O'Rourke*[a], *Theodore Reich*[a, c]

Departments of [a]Psychiatry, [b]Biostatistics and [c]Genetics, Washington University School of Medicine and the Jewish Hospital of St. Louis, St. Louis, Mo., USA

Key Words. Ascertainment · Proband · Twin study · Probandwise concordance

Abstract. When families are ascertained through affected twins, as for example when twin probands are selected from a registry and their non-twin relatives studied, a correction for ascertainment bias is needed. It is shown that probandwise counting (where relatives of doubly ascertained twin pairs are counted twice) is the appropriate method. The bias resulting from pairwise counting is given and depends on the genetic model and on the probability of selecting an affected twin as a proband. For the multifactorial and generalized single major locus models the bias is small, and the problems associated with nonindependent ascertainment are negligible in practice.

Introduction

Twin studies provide a unique data base for the investigation of human diseases, and existing twin registries have proven a valuable source of clinical data for disorders ranging from schizophrenia [*Gottesman and Shields,* 1972] to diabetes [*Harvald and Hauge,* 1976; *Pyke and Nelson,* 1976] and breast cancer [*Holm* et al., 1980]. However, if observations are limited solely to MZ and DZ pairs, restrictive environmental assumptions must be made since only two unknown model parameters may be estimated from two observations. Recently, much attention has been given to genetic models which allow for phenomena such as the transmission of environmental factors from parent to offspring (cultural transmission) and effects due to a contemporaneous environment of rearing [*Wright,* 1931; *Rao* et al., 1976; *Eaves,* 1976; *Cloninger* et al., 1979; *Rice* et al., 1980]. Accordingly, the range of hypotheses testable from twin data can be broadened by including, for example, the parents and siblings of the twins.

Essen-Möller and Fischer [1979] point out, however, that there is confusion in the literature on the appropriate procedure for taking the method of ascertainment into account when families are sampled through

[1] Supported in part by USPHS grants AA-03539, GM-28067, MH-25430, MH-31302 and MH-14677.

affected twins. To estimate the prevalence of affected non-twin relatives in an ascertained sample, they discuss pairwise counting (where each relative is counted once), probandwise counting (where relatives of doubly ascertained twins are counted twice), and argue that both approaches may be inadequate since affected non-twin relatives cannot themselves be probands. In reporting rates in non-twin relatives of ascertained twins, *Kringlen* [1967] used pairwise rates, whereas other investigators [cf. *Gottesman and Shields,* 1972; *Fischer,* 1973] report probandwise rates, although without a formal justification [cf. *Allen* et al., 1967]. In what follows we show that the probandwise rate is indeed the correct one, and we quantify the degree of bias resulting from the use of the pairwise rate.

Table I. Summary of notations used

$K_{P\text{-twin}}$	Prevalence of illness in twins
$K_{P\text{-rel}}$	Prevalence of illness in relatives (e.g., parents, siblings, etc.)
$K_{R\text{-twin}}$	Probability that the co-twin of an affected twin is affected
$K_{R\text{-rel}}$	Probability that a relative of an affected twin is affected
π	Probability that an affected twin is a proband
X	Number of affected twins in a twin pair, with X = 0, 1 or 2
Y	Number of probands in a twin pair, with $Y \leq X$
Prob ( )	Probability of the event in parentheses
Prob (asc)	Probability of ascertaining a twin pair
r	Correlation in liability for the multifactorial model
q, f_1, f_2, f_3	Parameters of the single major locus model with two alleles A, a, with q the frequency of a and f_1, f_2, f_3 the probability that an individual with genotype *AA, Aa, aa,* respectively, is affected

Methods

Estimation of the Prevalence of Affected Co-Twins

Let us first review the need for the use of probandwise concordance when calculating the prevalence of illness in co-twins of affected twins sampled using the proband method, and then generalize to the prevalence of illness in the relatives of the twins. The correction for ascertainment bias was developed by *Weinberg* [1928] and has been discussed in detail by *Morton* [1959], *Crow* [1965], *Allen* et al. [1967], *Gottesman and Shields* [1972] and *Smith* [1974].

Consider first the case when the twins are ordered (T_1, T_2), say by birth order. Let A and U denote affected and unaffected, respectively, and p_{AA}, p_{AU}, p_{UA}, p_{UU} denote the probabilities of the various affectational possibilities. The prevalence of the disease in twins, assuming $p_{AU} = p_{UA}$ is $K_{P\text{-twin}} = p_{AA} + p_{AU}$ and we wish to estimate $K_{R\text{-twin}} = p_{AA}/(p_{AA} + p_{AU})$.

Let π denote the probability that an affected twin will be a proband in the study. Then, using the notation summarized in table I,

$$\text{Prob (asc}|X = 2) = \pi^2 + 2\pi(1-\pi) = \pi(2-\pi), \quad (1)$$

$$\text{Prob (asc}|X = 1) = \pi, \quad (2)$$

$$\text{Prob (asc}|X = 0) = 0, \quad (3)$$

$$\text{Prob (asc)} = \pi(2-\pi)p_{AA} + 2\pi p_{AU}. \quad (4)$$

That is, the probability of ascertaining a pair conditional on having both members affected is the probability that they are both probands (π^2) plus the probability that exactly one is a proband; the probability of ascertaining a pair conditional on having exactly one affected member is π; the probability of ascertaining a pair both of whom are unaffected is 0. Therefore, the probability of ascertaining a pair is the weighted sum given in equation 4.

Let P_1 denote the probability in the ascertained sample that both twins will be affected and one is a proband, P_2 the probability that both will be affected and both are probands, and P_3 the probability that one is affected and a proband.

Then

$$P_1 = \text{Prob}\,(X=2, Y=1|\text{asc}) = 2\pi(1-\pi)p_{AA}/\text{Prob(asc)}, \quad (5)$$

$$P_2 = \text{Prob}\,(X=2, Y=2|\text{asc}) = \pi^2 p_{AA}/\text{Prob(asc)}, \quad (6)$$

$$P_3 = \text{Prob}\,(X=1, Y=1|\text{asc}) = 2\pi p_{AU}/\text{Prob(asc)}, \quad (7)$$

The P_i's are the expected proportions for the quantities actually observed in the sample, and we note that

$$\frac{2P_2 + P_1}{2P_2 + P_1 + P_3} = \frac{p_{AA}}{p_{AA} + p_{AU}} = K_{R\text{-twin}}, \quad (8)$$

so that an asymptotically unbiased estimate of $K_{R\text{-twin}}$ may be obtained by counting each doubly ascertained pair twice. This is remarkable both in the fact that this is true independently of the genetic/environmental model and independently of the value of π. In contrast the pairwise rate is given by

$$\frac{P_1 + P_2}{P_1 + P_2 + P_3} = \left\{\frac{(2-\pi)(p_{AA} + p_{AU})}{2(p_{AA} + p_{AU}) - \pi p_{AA}}\right\} K_{R\text{-twin}}, \quad (9)$$

which depends both on the values of p_{AA} and p_{AU} and on π. Note, however, that as π becomes small, then the term in brackets approaches 1.

Estimation of the Prevalence of Affected Relatives

Now consider the case where twin pairs are sampled using the proband method and that a particular type of relative, e.g., parent or sibling, is studied. Let (T_1, T_2, S) denote an ordered trio consisting of twin 1, twin 2, and relative, and let p_{ijk} denote the probability in a random sample that T_1, T_2 and S have phenotypes i, j, k, respectively. Thus, p_{AAA} is the probability that all three are affected, p_{AAU} the probability that both twins are affected and the relative unaffected, etc. We use a '·' to indicate a marginal distribution, so that $p_{A\cdot\cdot} = K_{P\text{-twin}}$, $p_{\cdot\cdot A} = K_{P\text{-rel}}$, $p_{AA\cdot} = p_{AA}$ (as above), and $p_{A\cdot A}$ the probability that twin 1 and the relative are affected.

Since ascertainment is limited to the twins T_1 or T_2, the probability of ascertainment, Prob(asc), is the same as that given in equation 4,

$$\text{Prob(asc)} = \pi(2-\pi)p_{AA\cdot} + 2\pi p_{AU\cdot} \quad (10.1)$$

$$= 2\pi p_{A\cdot\cdot} - \pi^2 p_{AA\cdot} \quad (10.2)$$

$$= 2\pi K_{P\text{-twin}} - \pi^2 K_{P\text{-twin}} K_{R\text{-twin}}. \quad (10.3)$$

The probabilities in the ascertained sample are

$$Q_1 = \text{Prob}[(A,A,A), Y = 2|\text{asc}] = \pi^2 p_{AAA}/\text{Prob(asc)} \quad (11)$$

$$Q_2 = \text{Prob}[(A,A,U), Y = 2|\text{asc}] = \pi^2 p_{AAU}/\text{Prob(asc)} \quad (12)$$

$$Q_3 = \text{Prob}[(A,A,A), Y = 1|\text{asc}] = 2\pi(1-\pi)p_{AAA}/\text{Prob(asc)} \quad (13)$$

$$Q_4 = \text{Prob}[(A,A,U), Y = 1|\text{asc}] = 2\pi(1-\pi)p_{AAU}/\text{Prob(asc)} \quad (14)$$

$$Q_5 = \text{Prob}[(A,U,A), Y = 1|\text{asc}] = \pi p_{AUA}/\text{Prob(asc)} \quad (15)$$

$$Q_6 = \text{Prob}[(A,U,U), Y = 1|\text{asc}] = \pi p_{AUU}/\text{Prob(asc)} \quad (16)$$

$$Q_7 = \text{Prob}[(U,A,A), Y = 1|\text{asc}] = \pi p_{UAA}/\text{Prob(asc)} \quad (17)$$

$$Q_8 = \text{Prob}[(U,A,U), Y = 1|\text{asc}] = \pi p_{UAU}/\text{Prob(asc)}, \quad (18)$$

so that the probandwise rate in the relatives is given by

$$\frac{2Q_1 + Q_3 + Q_5 + Q_7}{2Q_1 + 2Q_2 + Q_3 + Q_4 + Q_5 + Q_6 + Q_7 + Q_8} = \frac{p_{A\cdot A}}{p_{A\cdot\cdot}} = K_{R\text{-rel}}, \quad (19)$$

Equation 19 follows in a straightforward way using the formulas $p_{AUA} = p_{A\cdot A} - p_{AAA}$ and $p_{AU\cdot} = p_{A\cdot\cdot} - p_{AA\cdot}$ to simplify the expressions in the equation.

The pairwise rate α is seen to be

$$\alpha = \frac{Q_1 + Q_3 + Q_5 + Q_7}{\text{Prob(asc)}} \quad (20.1)$$

$$= \frac{2K_{P\text{-twin}} K_{R\text{-rel}} - \pi p_{AAA}}{2K_{P\text{-twin}} - \pi K_{P\text{-twin}} K_{R\text{-rel}}} \quad (20.2)$$

which depends on the parameter π and on the joint probability that all three individuals are affected.

The Bias Resulting from Use of Pairwise Rates

From equation 20.2, we see that the bias from the use of the pairwise rate approaches 0 as π approaches 0. This is clear since as π becomes

Table II. Value of the pairwise rate α as compared to $K_{R\text{-}rel}$ for the multifactorial model

K_p	r	$K_{R\text{-}rel}$	π	α
0.30	0.7	0.637	1.00	0.583
			0.75	0.600
			0.50	0.615
			0.25	0.627
	0.5	0.523	1.00	0.480
			0.75	0.494
			0.50	0.505
			0.25	0.515
0.10	0.7	0.470	1.00	0.418
			0.75	0.434
			0.50	0.448
			0.25	0.459
	0.5	0.320	1.00	0.286
			0.75	0.295
			0.50	0.304
			0.25	0.313
0.01	0.7	0.266	1.00	0.233
			0.75	0.244
			0.50	0.251
			0.25	0.259
	0.5	0.129	1.00	0.118
			0.75	0.121
			0.50	0.124
			0.25	0.126

Table III. Value of the pairwise rate α as compared to $K_{R\text{-}rel}$ for the generalized single major locus model

q	f_1	f_2	f_3	K_p	K_R	π	α
0.0513	0.0	1.0	1.0	0.10	0.544	1.00	0.533
						0.75	0.537
						0.50	0.540
						0.25	0.542
0.0050	0.0	1.0	1.0	0.01	0.504	1.00	0.503
						0.75	0.504
						0.50	0.504
						0.25	0.504
0.3162	0.0	0.0	1.0	0.10	0.433	1.00	0.401
						0.75	0.411
						0.50	0.419
						0.25	0.423
0.1000	0.0	0.0	1.0	0.01	0.303	1.00	0.294
						0.75	0.297
						0.50	0.299
						0.25	0.301
0.1000	0.0	0.5	1.0	0.10	0.325	1.00	0.315
						0.75	0.318
						0.50	0.320
						0.25	0.323
0.0100	0.0	0.5	1.0	0.01	0.258	1.00	0.257
						0.75	0.257
						0.50	0.257
						0.25	0.257

small, there will be few twin pairs with two probands, with the ratio of doubly ascertained to singly ascertained pairs being $\pi/(2\text{-}2\pi)$ in pairs where both twins are affected. To assess the bias when π is large, it is necessary to consider particular models of disease transmission in order to obtain the joint probabilities needed for computation of the expected value of the pairwise rate. We will consider the multifactorial model and the generalized single major locus model.

For the multifactorial model, $K_{R\text{-}rel}$ depends on the prevalences in twins and in the relatives and on the correlation in liability between the twins and between a twin and his relative [*Falconer*, 1965; *Smith*, 1970; *Reich* et al., 1972; *Curnow and Smith*, 1975]. In table II we assume for simplicity that $K_{P\text{-}twin} = K_{P\text{-}rel}$ and that the twin-twin and twin-relative correlations are equal to a common r, and show the value of $K_{R\text{-}twin}$ along with the α of formula 20. The bias decreases as r decreases, and we display results for r of 0.7 and 0.5. In the case of polygenic transmission and random mating, a parent-offspring correlation of 0.5 would correspond to a trait with 100% heritability. The probabilities were computed using the methods described in *Rice* et al. [1980].

For the generalized single major locus model [*Suarez* et al., 1976], we consider a locus with two alleles A,a and genotypes *AA, Aa, aa*. The model is parameterized in terms of q, the gene frequency of allele a, and the penetrances f_1, f_2, f_3, the probabilities that individuals with genotypes *AA, Aa, aa,* respectively, are affected. Table III indicates the bias for dominant, recessive, and additive models, where the relative is a sibling and where the twins are DZ.

The above analysis shows that probandwise rates are the correct ones when the relativies of ascertained twins are studied. However, tables II and III indicate that in practice the pairwise and probandwise rates will be close to one another, even with large π, and even at the 'extremes' of the two models. This is consistent with the various studies which have calculated both types of rates and report very little difference.

Discussion

The above analysis was motivated by the twin data on schizophrenia where the rates of affected parents and siblings of twins have been reported to be lower than those reported in studies where probands are not twins [*Essen-Möller and Fischer,* 1979]. We have shown that this difference is not due to problems associated with ascertainment. These authors point out that these differences may also be due to less thorough study of the non-twin relatives or be simply due to 'chance' as suggested by *Gottesman and Shields* [1976]. Although environmental effects (prenatal and postnatal) are no doubt more similar for DZ siblings than for singleton siblings, it is not yet clear whether or not they are relevant to the pathogenesis of schizophrenia.

The corrections here for ascertainment are simply those one would use if sibships were sampled but probands were restricted, say, to the first- or second-born. However, they are in general more important in twin research.

For example, *Fischer* [1973] drew her sample from a Danish central register of psychiatric hospital admissions, matched against a national twin birth cohort register [*Hauge,* 1981] so that π is nearly 1.

This result indirectly addresses the problems associated with nonindependent ascertainment of probands. The formulas derived above assume that in a doubly affected twin pair, the probability that an affected twin will be a proband conditional on his co-twin being a proband is π. It is possible, of course, that this conditional probability is higher due to a greater awareness of the illness once the first twin case falls ill or due to referral to the same familiar treatment facility. The problems with secondary ascertainment have been considered by *Smith* [1974] and *Allen and Hrubec* [1979]. Nonetheless, the mathematically precise corrections should, in general, lie between the probandwise and pairwise rates, and the above results indicate that these differences are trivial when considering the morbid risks in the relatives of twins.

References

Allen, G.; Harvald, B.; Shields, J.: Measures of twin concordance. Acta genet. *17:* 475–481 (1967).

Allen, G.; Hrubec, Z.: Twin concordance; a more general model. Acta Genet. med. Gemell. *28:* 3–13 (1979).

Cloninger, C.R.; Rice, J.; Reich, T.: Multifactorial inheritance with cultural transmission and assortative mating. II. A general model of combined polygenic and cultural inheritance. Am. J. hum. Genet. *31:* 176–198 (1979).

Crow, J.F.: Problems of ascertainment in the analysis of family data: In Neel, Shaw, Schull,

Genetics and the epidemiology of chronic diseases (Washington 1965).

Curnow, R.N.; Smith, C.: Multifactorial models for familial diseases in man. J. R. Stat. Soc. A. *138:* 131–169 (1975).

Eaves, L.J.: The effect of cultural transmission on continuous variation. Heredity *37:* 41–57 (1976).

Essen-Möller, E.; Fischer, M.: Do the partners of dizygotic schizophrenic twins run a greater risk of schizophrenia than ordinary siblings? Hum. Hered. *29:* 161–165 (1979).

Falconer, D.S.: The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann. hum. Genet. *29:* 51–76 (1965).

Fischer, M.: A Danish twin study of schizoprenia. Acta psychiat. scand., suppl. 238 (1973).

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

Gottesman, I.I.; Shields, J.: A critical review of recent adoption, twin, and family studies of schizophrenia: behavioral genetics perspectives. Schiz. Bull. *2:* 360–401 (1976).

Harvald, B.; Hauge, M.: Diabetes in twins. Nord. Council Arct. Med. Res. Rep. *13:* 13–15 (1976).

Hauge, M.: The Danish twin register; in Mednick, Baert, Prospective longitudinal research in Europe: an empirical basis for the primary prevention of psychosocial disorders (Oxford University Press, Oxford 1981).

Holm, N.V.; Hauge, M.; Harvald, B.: Etiologic factors of breast cancer elucidated by a study of unselected twins. J. natn. Cancer Inst. *65:* 285–298 (1980).

Kringlen, E.: Heredity and environment in the functional psychoses (Heinemann, London 1967).

Morton, N.E.: Genetic tests under incomplete ascertainment. Am. J. hum. Genet. *11:* 1–16 (1959).

Pyke, D.A.; Nelson, P.G.: Diabetes mellitus in identical twins; in Creutzfeldt, Köbberling, Neel, The genetics of diabetes mellitus (Springer, New York 1976).

Rao, D.C.; Morton, N.E.; Yee, S.: Resolution of cultural and biological inheritance by path analysis. Am. J. hum. Genet. *28:* 228–242 (1976).

Reich, T.; James, J.W.; Morris, C.A.: The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. Ann. hum. Genet. *36:* 163–184 (1972).

Rice, J.; Cloninger, C.R.; Reich, T.: Analysis of behavioral traits in the presence of cultural transmission and assortative mating: applications to IQ and SES. Behav. Genet. *10:* 73–92 (1980).

Smith, C.: Heritability of liability and concordance in monozygous twins. Ann. hum. Genet. *34:* 85–91 (1970).

Smith, C.: Concordance in twins: methods and interpretation. Am. J. hum. Genet. *26:* 454–466 (1974).

Suarez, B.K.; Reich, T.; Trost, J.: Limits of the general two-allele single locus model with incomplete penetrance. Ann. hum. Genet. *40:* 231–244 (1976).

Weinberg, W.: Mathematische Grundlage der Probandenmethod. Z. indukt. Abstamm.-Vererb.-Lehre *48:* 179–338 (1928).

Wright, S.: Statistical methods in biology. Proc. Am. Stat. Ass. *26:* 155–163 (1931).

John P. Rice, PhD,
Department of Psychiatry,
Washington University School of Medicine,
and Jewish Hospital of St. Louis,
216 South Kingshighway,
St. Louis, MO 63110 (USA)