Department of Zoology University of California (U.S.A.)

# The Ratio of monozygotic to dizygotic affected twins and the frequencies of affected twins in unselected data

**Kurt Stern**

The significance of comparisons between the frequencies of monozygotic and dizygotic twins concordant for specific traits depends on the methods of ascertainment. With complete ascertainment in populations with twin material unselected except for the presence of a given trait, it has been assumed (1) that the ratio of monozygotic to dizygotic pairs is that of the general population and (2) that the relative frequency of twin pairs to singly born individuals is also like that in the general population. It will be shown that these assumptions are not valid.

Let *m* be the frequency of a certain genotype among the offspring of « suitable » parents, i.e. those which are genetically able to produce the genotype. Let *n* be the penetrance with which the « affectable » genotype expresses itself in a specific phenotype.

Then, among all *monozygotic twin pairs* from suitable parents the frequency of affectable twins is *m*, since monozygotic pairs represent a single zygote. The only phenotypically ascertainable pairs are the positively concordant, whose frequency is

$$mn^2$$

and the discordant ones whose frequency is

$$2mn(1—n)$$

The sum of the frequencies of these two ascertainable monozygotic twin pairs is

$$mn(2—n) \qquad (1)$$

Among *dizygotic twin pairs* from suitable parents the frequency of pairs (a) both partners of which are affectable is $m^2$ and of pairs (b) only one of whose partners is affectable is $2m(1—m)$. Among pairs (a) the frequency of positively concordant pairs is

$$m^2n^2$$

and of discordant pairs

$$2m^2n(1—n)$$

Ricevuto il 22. II. 1958.

Among the pairs (b) no positively concordant ones are possible. The frequency of discordant pairs in this group is

$$2mn(1—m)$$

The sum of the frequencies of the three ascertainable dizygotic twin pairs is

$$mn(2—mn) \qquad (2)$$

If the ratio of monozygotic to dizygotic twin pairs in the general population is $k$, then the ratio, k′, of the two zygotic twin pair types ascertainable for a specific trait is $k$ times the ratio formed from terms (1) and (2), namely

$$k' \left(\frac{\text{ascertainable monozygotic pairs}}{\text{ascertainable dizygotic pairs}}\right) = k \frac{2—n}{2—mn} \qquad (3)$$

For all positive values of m and n the term $(2—n)/(2—mn)$ is smaller than 1. It does not fall below 0.5. Depending on the frequency m of the affectable genotype from suitable parents and the degree n of penetrance, *the expected ratio*, k′, *of ascertainable monozygotic to dizygotic twin pairs lies thus between* ½ *and* 1 *of the ratio*, $k$, *of monozygotic to dizygotic pairs in the general population.*

This deviation from the ratio of the two types of twins in the general population is due to two independent phenomena resulting in opposite trends. One of these is an overrepresentation of ascertained dizygotic pairs by uncorrected inclusion of those in which one twin does not belong to the affectable genotype in contrast to the ascertained monozygotic pairs in which each twin is affectable. The other is an underrepresentation of dizygotic pairs due to incomplete penetrance which results in a greater loss of dizygotic than of monozygotic pairs. Specifically, the loss of dizygotic pairs consisting of the sum of negatively concordant among those pairs, both partners of which are affectable and those where only one partner is affectable is

$$m^2(1—n)^2+2m(1—m)\ (1—n) \qquad (4)$$

For monozygotic pairs the loss is

$$m(1—n)^2 \qquad (5)$$

For $m=1$, the losses of the two types of twins due to non-penetrance are equally $(1—n)^2$ but for all values of $m<1$, expression (4) is larger than (5), i.e. the losses of dizygotic pairs are greater than those of monozygotic.

When twin pairs have not been individually diagnosed as either monozygotic or dizygotic it is customary to derive the ratio of the two types by means of Weinberg's differential method. Assuming equality of the sexes, the ratio, h, of like-sexed to unlike-sexed twin pairs in the general population, is

$$h = \frac{MZ + \frac{1}{2}\,DZ}{\frac{1}{2}\,DZ}$$

which is equal to

$$h = \frac{2\ MZ}{DZ} + 1$$

It follows that the ratio, h′, of ascertainable affected like-sexed to unlike-sexed twins pairs is

$$h' \left( \frac{\text{ascertainable like-sexed twin pairs}}{\text{ascertainable unlike-sexed twin pairs}} \right) = 2k' + 1 \qquad (6)$$

In various studies of the expectancy of specific phenotypes in cotwins of affected partners, the ratio of ascertained monozygotic to dizygotic pairs deviates considerably from the ratio in the general population. Whenever the monozygotic twins are significantly more frequent than in the general population, selective ascertainment must be involved. Significant deviations in the opposite direction, an underrepresentation of monozygotic twins, may at least partly be due to the phenomenon expressed by (3), that is due to the fact that in unselected series the ratio of ascertainable affectable monozygotic to dizygotic twins usually should be lower than that of such twins in general.

The twin data on schizophrenia tabulated by Kallmann (1953) from the works of five different investigators may serve as an illustration. In each set of data the frequency of monozygotic pairs is less than one half of that of dizygotic pairs (the ratio k in the general populations being close to 1:2). The combined ratio from the five studies, 378:985, as well as the ratio from the largest single study alone, 268:685, deviates significantly from k. These deviations may possibly be explained by the differential ascertainability of affected twin pairs.

It should be made explicit that the foregoing treatment assumes not only a specific penetrance but also that it must be applied separately to different possible genotypes of « suitable » parents. This may be shown by the simple example of a recessive, fully penetrant genotype *aa* such as the one causing albinism. The affectable individuals aa can come from three different types of suitable parents (1) Aa×Aa, (2) Aa×aa, and (3) aa×aa. In all three cases n= 1, but m, the frequency of the affectable genotypes in the offspring is 0.25, 0.5 and 1.0 respectively. Entering these values into formula (3) k′ becomes equal to 4/7 k for mating (1), 2/3 k for mating (2), and k for mating (3). In the case of a rare trait like albinism the overwhelming number of affected individuals comes from mating (1). Therefore, the ratio of monozygotic to dizygotic albino twin pairs will be only slightly larger than 4/7 of the ratio of these twin types in the general population.

For a trait dependent on homozygosity of a frequent allele all three types of matings with their different m values contribute appreciably to the mean ratio k′ observed in twin pairs — ascertained by belonging to the specified trait. Thus, in a population in which the frequency of the blood group allele $M^N$ is 0.5, the three matings capable of producing the $M^N$ $M^N$ genotype occur in the proportions 4:4:1 and the weighted mean of the three k′ ratios is 125/189= 0.66. Monozygotic twin pairs ascertained by their belonging to blood group N should thus be only 0.66 times as frequent in proportion to dizygotic twin pairs with one of two N partners, as the ratio of the two twin types in the general population.

## II.

Similar to the loss of concordant nonpenetrant twin pairs due to lack of ascertainability which differentially affects monozygotic and dizygotic twins, is the differential loss of nonpenetrant single-born individuals versus twin pairs. The loss of single-born among the affectable is

$$m(1-n) \tag{7}$$

If f=number of all twin pairs in the general population expressed as a portion of all individuals in the population and, as before, k=frequency of monozygotic pairs among twin pairs, then the frequency of ascertainable twin pairs is

$$(MZ+DZ) = f\{(k[m-m(1-n^2)]+(1-k)[m^2+2m(1-m)-m^2(1-n)^2-2m(1-m)(1-n)]\}$$

which simplifies to $(MZ+DZ) = fmn\,[2-n\,(k+m-km)]$

The total frequency of ascertainable single-born is

$$S = (1-f)mn$$

Thus the fraction of twin pairs among all ascertainable individuals is

$$d = \frac{MZ + DZ}{MZ + DZ + S} = \frac{f[2 - n(k + m-km)]}{f[2 - n(k + m-km)] + 1-f} \tag{8}$$

Since f and k are both less than 1, this expression is always greater than f if either m or n are smaller than 1. The maximal value for d is 2f.

The fact that d>f may also be seen by considering the earlier terms (4), (5) and (7) for the losses due to non ascertainment of monozygotic, dizygotic and single-born individuals. Regarding m, the loss of single-born is solely proportional to it while the losses of the twin pairs are partly proportional to m, and partly only to $m^2$ and $m(1-m)$. Thus, the losses of twin pairs are relatively less than those of single-born individuals. Likewise, regarding n, the loss of single-born is solely proportional to $(1-n)$ while those of the twin pairs are partly proportional to $(1-n)$ and partly only to $(1-n)^2$. Thus, again, the losses of twin pairs are relatively less than those of single-born individuals.

Consequently, the numerator of (8), which is equal to (MZ+DZ), decreases less with decreasing values of m and n than the denominator which is equal to (MZ+ +DZ+S). It follows that in unselected populations of affected individuals the *expected frequency of ascertainable twin pairs is greater than in the general population.* The increase is by a factor which lies between 1 and 2.

Three theoretical examples may indicate the magnitude of the effect if reasonable numerical values are substituted for f, k, m and n, namely (1) f = 0.01, k = 0.33, m = 0.25 and n = 1, (2) f, k, and m as before, n = 0.5 and (3) f, k, and m as before n = 0.02. With (1) d = 0.0150, with (2) d = 0.0174 and with (3) d = 0.0197. Comparison with f = 0.01 shows that the expected frequency of twins in the ascer-

tained sample of the affectable population is 50, 74 and 97 percent higher, respectively, than in the general population.

Actual values for d, from seven different studies on affected individuals in adult German populations as listed by Luxemburger (1940) vary from 0.0055 to 0.0156 with a mean of 0.0095 as compared to an f = 0.0090. The more recent work of Herndon and Jennings (1951) on polyomyelitis in North Carolina yielded a value uncorrected for age, of d = 0.01440 + 0.00191 as compared to the value for the general United States population of f = 0.01161. The difference between d and f is not significant though in the expected direction.

In his 'Comments on the analysis of twin samples', Gordon Allen (1955) has already raised some of the issues which form the subject of the two preceding sections. He pointed out that twin data should best be expressed in terms of index cases instead of twin pairs. In a large and representative sample, the number of twin index cases will be proportional to the frequency of the respective category in all twins. On the contrary, the distributions of pairs of twins are distorted in relation to those of the general population. Allen also paid particular attention to the effects of sample size on concordance rates.

## III

Theoretically, unselected twin data permit an evaluation of *m*, the mean frequency of the affectable genotype among the offspring of suitable parents. It follows from equation (3) that

$$m = \frac{2k' - k(2 - n)}{k'n} \tag{9}$$

The values of k′ and k are known from the ascertained and the general population respectively. The penetrance, n, can be derived (after Diehl and v. Verschuer, 1932, Luxenburger, 1940 and others) from the concordance frequency, e, of affected monozygotic twins:

$$n = \frac{2e}{1+e} \tag{10}$$

For any specific trait, the value for m obtained from equation (9) can be compared with the value of m derived as is usually done from the concordance rates of monozygotic and dizygotic twins.

Both values of m depend on terms which have large errors. This brings it about that equation (9) sometimes cannot be applied to empirical data. When values of *m* can be obtained, they are of interest since they may be compared with expectations for *m* under various assumptions concerning the genotypes of affected individuals. However, the use of formula (9) is correct only if the value of m is the same for the offspring of all suitable matings, i.e. if all pairs of suitable parents have the same

genotypes, and if all affectable offspring are genetically alike. In case of a hypothesis of recessiveness of a rare trait the great majority of suitable matings would be Aa×Aa with an expected m = 0.25. In case of a hypothesis of dominant inheritance most affectable individuals would come from Dd×dd matings with an expected m = 0.5.

It should be emphasized that these considerations oversimplify a variety of aspects of interaction of genetic and non-genetic factors. Thus, the derivation of the degree of penetrance, n, from the frequency of concordance in monozygotic twin pairs undoubtedly produces a value which is too high if applied to individuals, such as dizygotic twins who are not isogenic even though the assumed homozygous recessive or the dominant main genes may be present in both of them.

The derivation of genetic hypotheses from concordance frequencies of twins has been treated extensively by Huizinga and v.d. Heiden (1957) whose work appeared in print after the present paper had been completed. In manner of treatment of their different though related topics the two studies are intrinsically similar.

**Acknowledgement**

I am grateful to Professors E.R. Dempster, Berkeley and O.v. Verschuer, Münster for critical reviews of the manuscript.

**Summary**

With complete ascertainment, in populations with twin material unselected except for the presence of a specific trait (1) the ratio of ascertainable monozygotic to dizygotic twin pairs lies between ½ and 1 of the ratio in the general population and (2) the ratio of all ascertainable twin pairs to the total number of individuals lies between one and two times of that found in the general population. Formulae are derived which relate the frequencies of ascertainable twin pairs to the frequencies, from « suitable » parents, of the « affectable » genotype and to its penetrance. The calculated expectations are compared with some observed values.

## Literature

ALLEN, G. 1955: Comments on the analysis of twin samples. Acta Geneticae Medicae et Gemellologiae, 4: 143-160.

DIEHL, K. and O. v. VERSCHUER, 1933: Zwillingstuberkulose 500 p., Jena: Fischer.

HERNDON, C. N. and R. G. JENNINGS, 1951: A twin-family study of susceptibility to polyomyelitis, Amer. J. Hum. Genet., *3*: 17-46.

KALLMANN, F. J., 1953: Heredity in health and mental disorder. 315 p. New York: Norton and Co.

LUXENBURGER, H., 1940: Die Zwillingsforschung als Methode der Erbforschung beim Menschen. In Hdbch. d. Erbbiol. d. Menschen. (G. Just., edit.) *2*: 213-248.

RIASSUNTO

Con l'accertamento totale, in popolazioni comprendenti materiale gemellare non selezionato se non per la presenza di un carattere specifico, si trova che:

1) La proporzione delle coppie accertabili monozigotiche rispetto alle dizigotiche varia fra ½ e 1 in rapporto alla proporzione nella popolazione in generale:

2) la proporzione del totale delle coppie gemellari accertabili rispetto al numero totale di individui varia fra una e due volte quella trovata nella popolazione in generale.

Vengono derivate alcune formule che mettono in rapporto le frequenze delle coppie gemellari accertabili con le frequenze, derivanti da genitori « adatti », del genotipo « colpibile » e della sua penetranza. Le previsioni calcolate vengono raffrontate con alcuni valori osservati.

RÉSUMÉ

Après vérification complète, sur des populations comprenant du matériel de jumeaux non-sélectionné sauf pour la présence d'un caractère spécifique, on trouve que:

1) la proportion entre les couples de jumeaux monozygotiques et dizygotiques vérifiables varie entre ½ et 1 par rapport à la proportion trouvée dans la population générale;

2) la proportion entre le nombre total de couples vérifiables et le nombre total d'individus varie entre une et deux fois la proportion trouvée dans la population générale.

On dérive quelques formules mettant en rapport les fréquences des couples de jumeaux vérifiables avec les fréquences, par des parents « appropriés », du génotype « affectable » et de sa pénétrance. Les prévisions calculées sont comparées avec quelques valeurs observées.

## ZUSAMMENFASSUNG

Wen man eine Bevölkerung mit Zwillingsmaterial, das nur im Hinblick auf das Vorhandensein eines bestimmten Merkmals ausgelesen wird, insgesamt betrachtet, so findet man, dass:

1) gegenüber dem Verhältnis bei der Bevölkerung im Allgemeinen, dasjenige swischen den betreffenden eineiigen und zweieiigen Zwillingspaaren von ½ bis 1 schwankt;

2) das Verhältnis der insgesamt betroffenen Zwillingspaare zur Gesamtzahl der Individuen zwischen dem ein- zweifachen des in der Gesamtbevölkerung festgestellten Verhältnisses schwankt.

Daruus werden einige Formeln abgeleitet, in denen die Häufigkeit der in Frage kommenden Zwillingspaare den von « geeigneten » Eltern herrührenden Häufigkeiten des « anfälligen » Genotyps und dessen Penetration gegeübergestellt werden. Die daraus errechneten Dalem werden mit einigen beobachteten Werten verglichen.




https://doi.org/10.1017/S1120962300019326 Published online by Cambridge University Press