Acta Genet Med Gemellol 34:167-174 (1985)

Received 3 June 1985

Zygosity Partitioning of Small Twin Samples

Gordon Allen[1] **and Zdenek Hrubec**[2]
[1] *9326 West Parkhill Drive, Bethesda, Maryland, and* [2] *Radiation Epidemiology Branch, National Cancer Institute, Bethesda, Maryland*

Abstract. When the Weinberg estimate of the proportion of monozygotic pairs is quite deviant from that in the source population, it is likely to be wrong because Weinberg's difference is much less stable than the zygosity proportions. A formula is proposed for the probability distribution of possible compositions of a small sample of twins based on sex concordance in the sample and zygosity proportions in the source population.

Key words: Twin samples, Zygosity, Likelihood method, Weinberg's difference

INTRODUCTION

Weinberg's difference method [10] estimates the number of monozygotic (MZ) twins in a sample by subtracting the number of unlike-sexed pairs from the number of like-sexed pairs, assuming that half of the dizygotic (DZ) pairs are of unlike sex:

$$M = L - U \qquad D = 2U \tag{1}$$

Here M and D are estimated numbers of MZ and DZ pairs, respectively; L and U are observed numbers of like-sexed and unlike-sexed pairs.

The estimates are exactly right only when the DZ twins in the sample are equally divided between the sex-concordance types. Inequality of the types may result from several causes elsewhere discussed [1]:

a) The sex ratio in the population is not usually unity as Weinberg's method assumes; this error can almost always be neglected.
b) The twins may not be a random sample of the source population.
c) Sex may be correlated within DZ pairs; this possibility has been urged particularly by James [6] and if not completely incompatible with Weinberg's method, would require addition of a constant [1].
d) Sampling error is inevitable and in small samples may completely invalidate an estimate based on sex concordance.

The purpose of this paper is to explore a method of minimizing the impact of sampling variance on the estimated proportion of MZ pairs in a small series of twins under study. Since some information is usually available about twinning in the source popula-

https://doi.org/10.1017/S0001566000004682 Published online by Cambridge University Press

tion, use of that information can improve the estimate for a sample. The proposed calculations may repay the added labor when four conditions are met:

a) The sample is small; say, less than 35 pairs.
b) Zygosity cannot be diagnosed in all like-sexed pairs by direct methods; this is often the case in mortality and perinatal studies.
c) Important inferences require the best possible estimate of the proportion of MZ pairs.
d) The Weinberg estimate differs substantially from the proportions in the source population.

FORMULATION

The distribution of zygosity types in twin samples of specified size and sex composition combines two binomially distributed proportions. First, the proportion of MZ twins in a sample of N pairs, which is a function of the probabilities in the source population: m for MZ twins and d for DZ twins, where m + d = 1. Second, the proportion of sex-concordant pairs, c, among DZ twins, which has an expected value of virtually 0.5 for all ordinary sex ratios.

Let M be the number of MZ pairs in the sample and let D be the number of DZ pairs, where D_1 is the number of DZ like-sexed pairs and D_2 or U is the number of unlike-sexed pairs. L is the number of all like-sexed pairs. Then $N = U + L = M + D$ and $L = M + D_1$. Usually only U (= D_2) and L are observable.

If for a sample of size N the particular values for M, D_1, and L are X, Y, and Z respectively, then the likelihood of the sample is

$$\begin{aligned}\Pr(M = X, D_1 = Y|N, m, c) &= \Pr(M = X|N, m)\Pr(D_1 = Y|N, c, M = X)\\ &= P_M(X|N, m)\, P_{L\text{-}M}(Z - X|N - X, c)\\ &= \binom{N}{X} m^X (1 - m)^{N-X} \cdot \binom{N - X}{Z - X} c^{Z\text{-}X} (1 - c)^{N\text{-}Z} \qquad (2)\end{aligned}$$

where possible values of X range from 0 to Z. For an observed value of L = Z, the likelihood that M is the particular value, X, may be calculated for assumed parameter estimates of m and c. When c is 0.5, the second half of (2) becomes

$$\binom{N - X}{Z - X} 0.5^{N\text{-}X}$$

APPLICATION

To illustrate with an extreme but not improbable example, consider a small, randomly drawn sample of 11 pairs of twins consisting of 6 unlike-sexed and 5 like-sexed pairs. Weinberg's difference method gives an impossible result or implies no MZ pairs. However, if twinning is well documented in the source population (eg, for the United States, [5]), we can take the value of m from that population. We can then use the two binomial distributions to obtain the probability of each possible value of X, the number of MZ pairs in the sample.

If the proportion of MZ twinning were expected to be 25% (for convenience of illustration), the probability that X = 0 and that all 11 pairs are DZ is given by the last term of the binomial in the first half of formula (2), namely, $(0.75)^{11} = 0.0422$. The

https://doi.org/10.1017/S0001566000004682 Published online by Cambridge University Press

probability that any 11 DZ pairs would include 6 of unlike sex and 5 of like sex is obtained from the sixth term of the binomial in the second half of (2): $462 \times (0.5)^{11} = 0.2256$, if $c = 1/2$. The product of these probabilities, 0.0095, is the likelihood that a sample of 11 pairs of twins drawn randomly from this population will consist entirely of DZ pairs, including 6 of unlike sex.

A figure can be calculated similarly for the likelihood that the sample comprises 1 MZ and 10 DZ (6 of unlike and 4 of like sex), and for all other possible compositions of the sample, as shown in Table 1. Given that $L = 5$ and $N = 11$, the sum of all probabilities is 0.1225, with the most probable composition being 2 MZ and 9 DZ pairs. Among the six possible compositions, this has a relative probability of 0.35, and only a composition in which all 5 like-sexed pairs are MZ has a relative probability that would be excluded at the $p = 0.05$ level.

The method can just as well be applied in a case where Weinberg's rule yields a positive but unexpected estimate for the number of MZ pairs, with the prospect of improving the estimate. If the 11-pair sample comprises 5 unlike-sexed and 6 like-sexed pairs, Weinberg's method estimates 1 MZ pair, but the proposed method, as shown in Table 2, indicates that from this source population, in which $m = 0.25$, the most likely number is 2, and that 3 MZ pairs would be more probable than 1 pair.

ACCURACY OF THE PARAMETER ESTIMATE, m

Equation (2) assumes that m, the proportion of all twins that are MZ, is a true parameter of the source population, when in fact only an estimate is ever available, usually derived from birth statistics by the difference method. The estimate is subject to error not only because vital statistics are based on finite populations, but because the statistical source and the twin sample are likely to differ in the distribution of mothers by age, parity and race; variables which strongly affect the frequency of DZ twinning.

When m is based on a small source population, it may be unreliable and its variance tends to spread the distribution of likelihoods. To illustrate the order of magnitude of these errors, if the value of m used in the preceding example, 0.25, were derived from 1,500 twin deliveries, its standard error (using-Weinberg's formula given in the next section) would be one-tenth of the estimate, 0.025. The 95% confidence interval would be $0.25 \pm \pm 0.05$. In Table 3 the computations of Table 2 are repeated at these confidence limits, $m = 0.20$ and $m = 0.30$, and the results are displayed to the left and right of the set of likelihoods computed at $m = 0.25$. Since the direction of possible error in m is not known, any more elaborate treatment than formula (2) would necessarily combine the effects of positive and negative deviations, which tend to offset each other. The outcome can be crudely represented by averaging across the three columns in Table 3 to yield a new set of likelihoods, given in the last column. The averaged likelihoods are very similar to those in the second column based only on the central value, $m = 0.25$. Thus, the sampling variance of m is important if large, but has little effect on the spread.

More serious discrepancies in m between the twin sample and the source population can result when the twin sample is atypical in the maternal characteristics of age, parity, and race. If statistics of the source population define the effects of these variables on twinning, the estimate of m used in formula (2) can be adjusted by an inverse standardization. Like-sexed and unlike-sexed twins in subdivisions of the source population by age, parity, and race are weighted to match the composition of the sample, and summed.

https://doi.org/10.1017/S0001566000004682 Published online by Cambridge University Press

Weinberg's difference based on the weighted sums will yield a value of m appropriate for the twin sample.

INUTILITY OF THE STANDARD ERROR IN SMALL TWIN SAMPLES

Differences between two estimates, or deviations of an estimate from the expected value, are usually assessed by means of a standard error formula. Bulmer [3] provided standard error formulas for use with twin frequencies measured within a whole population including singletons, frequencies which are distributed approximately according to Poisson. Within a population of twins only, m and d add to 1 and are distributed binomially. To derive standard errors for such estimates, Weinberg [8] began with the variance in the form, pq/n, for the proportion of same-sexed pairs, s, and from that derived the variance of the zygosity estimates, in part as follows:

$$V(s) = V(1-s) = s(1-s)/N \qquad (3)$$

Variances of m and d are derived from that of s by use of the relations (1): $m = s - (1-s)$ and $d = 2(1-s)$. Their variances are four times the variance of s:

$$\begin{aligned} V(m) &= V[s-(1-s)] = V(2s) = 4V(s) = 4s(1-s)/N \\ V(d) &= V[2(1-s)] = 4V(s) = 4s(1-s)/N \end{aligned} \qquad (4)$$

To obtain an expression for the variance of the numbers M and D, instead of the proportions, m and d, one replaces m, s, and (1 – s) with M/N, L/N, and U/N:

$$\begin{aligned} V(M/N) &= (1/N^2)\ V(M) = 4\,LU/N^3 \\ V(M) &= 4\,LU/N = V(D) \end{aligned} \qquad (5)$$

Square roots of (4) and (5) are the desired standard errors, respectively:

$$SE(m) = 2\sqrt{s(1-s)/N} \qquad SE(M) = 2\sqrt{LU/N} \qquad (6,7)$$

Use of these standard error formulas assumes that the Weinberg estimate varies around the true value according to the normal distribution function. The binomial distribution approximates the normal distribution well enough for large numbers, so that formula (6) is suitable for evaluating the estimate of m among twins in the source population. The approximation is poor for small numbers, and further, when the two frequencies are very unequal as s and (1 – s) are likely to be, errors of estimation will be asymmetric around the parameter. In such cases the standard error formula may indicate only the rank order of certainty of Weinberg estimates in different twin samples.

In small twin samples the 95% confidence interval of Weinberg's difference, calculated from the above standard errors, is so wide that it may include very bad estimates of the proportion of MZ pairs. Consider three hypothetical samples of twins, displayed in Table 4. The first sample is that described in Table 1; the impossible Weinberg estimate of negative one MZ pair is only 1.1 standard error, by formula (7), from the population value of 25% MZ, quite acceptable by this criterion. If we used the t distribution or substituted N – 1 for N in formula (7), the Weinberg estimate would appear even more probable, when actually it is impossible.

A second sample of 11 pairs, consisting of 8 like-sexed and 3 unlike-sexed pairs, would yield a Weinberg estimate of 5 MZ, only 0.76 SE from 25%, while formula (2) would advise an estimate of 3 MZ pairs. A third sample, three times as large as the second

https://doi.org/10.1017/S0001566000004682 Published online by Cambridge University Press

Table 1. Absolute and relative probabilities of all possible compositions of a twin sample comprising 5 like-sexed and 6 unlike-sexed pairs, given that the expected proportion of MZ pairs is 0.25. Symbols are defined in the text.

Composition of the twin sample			Likelihood of M/D	Likelihood of D_1/D_2	Product of likelihoods; absolute probability	Relative probability
M	D_1	D_2				
0	5	6	0.042	0.226	0.0095	0.078
1	4	6	0.155	0.205	0.0318	0.259
2	3	6	0.258	0.164	0.0423	0.346
3	2	6	0.258	0.109	0.0282	0.230
4	1	6	0.172	0.055	0.0094	0.077
5	0	6	0.080	0.016	0.0013	0.010
					0.1225	1.000

Table 2. Absolute and relative probabilities of all possible compositions of a twin sample comprising 6 like-sexed and 5 unlike-sexed pairs, given that the expected proportion of MZ pairs is 0.25. Symbols are defined in the text.

Composition of the twin sample			Likelihood of M/D	Likelihood of D_1/D_2	Product of likelihoods; absolute probability	Relative probability
M	D_1	D_2				
0	6	5	0.042	0.226	0.0095	0.0467
1	5	5	0.155	0.246	0.0381	0.1866
2	4	5	0.258	0.246	0.0635	0.3111
3	3	5	0.258	0.219	0.0565	0.2765
4	2	5	0.172	0.164	0.0282	0.1382
5	1	5	0.080	0.094	0.0075	0.0368
6	0	5	0.027	0.031	0.0008	0.0041
					0.2042	1.0000

Table 3. Relative probabilities of all possible compositions of a twin sample comprising 6 like-sexed and 5 unlike-sexed pairs, as in Table 2, calculated for three different values of m, the expected proportion of MZ pairs, and the results averaged.

Composition of the twin sample: MZ pairs	Relative probabilities			Average
	m = 0.20	m = 0.25	m = 0.30	
0	0.0878	0.0467	0.0244	0.0530
1	0.2634	0.1866	0.1254	0.1918
2	0.3292	0.3111	0.2686	0.3030
3	0.2195	0.2765	0.3070	0.2677
4	0.0823	0.1382	0.1973	0.1393
5	0.0165	0.0368	0.0677	0.0403
6	0.0014	0.0041	0.0097	0.0051

and with the same proportions, gives a Weinberg difference of 15 MZ pairs; this is within 1.32 SE of 25%, but has a relative probability, by formula 2, of only 0.014. At the same time, the value 9.5 obtained by the likelihood method would be excluded by the Weinberg estimate ($P < 0.05$ in the right-hand column of Table 4).

Thus, estimates with modest deviations from the expected value of m, in terms of the standard error, may be quite unlikely, and capable of substantial improvement by the proposed likelihood calculations.

If the Weinberg estimate appears suspect and the labor of the likelihood method is unwarranted, an expedient is available. One may disregard the sex types of the twin pairs and assume zygosity proportions as in the source population. In the third example in Table 4, drawn from a population in which one-fourth of twin pairs are MZ, about 8 pairs are expected to be MZ. This figure is much closer to the best estimate of 9.5 than is the Weinberg difference, 15. Evidently, deviation of a Weinberg estimate from the expected frequency of m is more often due to sampling of sex concordance, magnified by the estimation procedure, than to sampling of zygosity proportions. In analyzing a particular sample, therefore, one should not use a deviant Weinberg estimate uncritically.

PARTLY DIAGNOSED SAMPLES

Often in studies of perinatal pathology some twin pairs are broken by death while others remain intact for study. Positive identification of some like-sexed pairs as MZ or DZ always enhances the value of the twin sample, but it should not be supposed that a few such determinations will assist in the overall partitioning of zygosity. A small number thus classified cannot be representative of the whole sample; they yield negligible information about the unclassified pairs. In conjunction with the proposed likelihood method, knowledge of some pairs will exclude extreme values, but the probability calculations are unchanged and the remaining possible compositions retain the same probabilities relative to one another.

For illustration, suppose that 3 pairs in the sample described in Table 1 were identified as 2 MZ and 1 DZ. This excludes lines 1, 2, and 6 of the table and leaves a choice among the three compositions given on lines 3, 4 and 5; two MZ pairs will still be the estimate of choice, with a relative probability of 0.53. Only rarely would partial classification of a sample alter the partitioning reached by formula (2), but in this example, if precise classification had identified 3 MZ pairs, the fourth composition would replace the third as the most probable.

APPLICATION TO TRAITS

An attractive, perhaps beguiling, feature of Weinberg's difference method, advanced by Weinberg, is that it can theoretically be applied directly to traits in a twin sample to estimate zygosity-specific frequencies, means, and even variances (but see cautions in [2] and [4]). A limitation is the implied assumption, often untenable, that all DZ pairs have the same value of whatever statistic is to be derived, regardless of sex.

The method is hardly applicable to small twin samples because the difference between like-sexed and unlike-sexed pairs, the primary criterion of statistical significance when Weinberg's method is used in this way, will usually be too small. However, when the likelihood method of formula (2) is used to partition the twins and particular traits are to

https://doi.org/10.1017/S0001566000004682 Published online by Cambridge University Press

Table 4. Comparison of relative probabilities computed by the likelihood method, using m = 0.25, with those computed from Weinberg's estimate and its standard error. Italicized relative probabilities indicate the best estimates by each method. Tails of the third distribution are not shown.

Twin sample structure			Relative probability by likelihood method	Weinberg estimate and SE	Relative probabilities based on Weinberg*
Observed		Postulated MZ pairs			
Like-sexed	Unlike-sexed				
5	6	0	0.078	−1±3.303	*0.277*
		1	0.259		0.242
		2	*0.346*		0.192
		3	0.230		0.140
		4	0.077		0.092
		5	0.010		0.056
8	3	0	0.017	5±2.954	0.038
		1	0.090		0.064
		2	0.209		0.095
		3	*0.279*		0.126
		4	0.232		0.149
		5	0.124		*0.159*
		6	0.041		0.149
		7	0.008		0.126
		8	0.001		0.095
24	9	6	0.056	15±5.117	0.017
		7	0.096		0.024
		8	0.136		0.032
		9	*0.161*		0.041
		10	*0.161*		0.050
		11	0.137		0.060
		12	0.099		0.069
		13	0.061		0.075
		14	0.032		0.080
		15	0.014		*0.081*
		16	0.005		0.080
		17	0.002		0.075
		18	0.000		0.069
		19			0.060
		20			0.050

* Relative values of those ordinates of the normal curve corresponding to possible compositions of the twin sample.

be reported, it may be appropriate to apply the same zygosity proportions to the traits with a caveat as to its significance. The method differs from Weinberg's only in substituting a weighted for a simple difference, as illustrated below.

On the assumption that T, the frequency or the average value of a trait, is the same in like-sexed and unlike-sexed DZ twins, and different in MZ twins,

$$\begin{aligned} MT_{mz} &= LT_l - D_l\ T_u \\ T_{mz} &= (LT_l - D_l\ T_u)/M \end{aligned} \tag{8}$$

where T_{mz}, T_l, and T_u are values of T for MZ, for all like-sexed, and for unlike-sexed twins, respectively; L is the observed number of like-sexed pairs, and M and D_l are estimated numbers of MZ and DZ pairs among L, as previously defined.

DISCUSSION

The illustrations have intentionally exaggerated the problem of *impossible* Weinberg estimates by using the low value of m = 0.25, to make bad estimates conspicuous. In contemporary Caucasian populations of usual maternal age distributions the figure is closer to 0.40, and an excess of unlike-sexed over like-sexed pairs is much less likely. At higher values of m, inaccuracy of the Weinberg estimate is more likely to escape notice, but the proposed likelihood calculations remain as advantageous as in the examples given.

Occasions to apply the proposed methods will occur when prior death or failure of contact prevents zygosity classification of some or all pairs and the Weinberg estimate appears to be deviant. The method of formula (2) is most clearly applicable to samples of less than 30 pairs. Partial application to larger samples for which Weinberg's estimate appears deviant may also be practical; formula (2) would then be applied not to all possible compositions of the sample as in Tables 1 and 2, but only to the value representing Weinberg's difference, to the proportion expected from the source population, and to numbers of MZ twins falling between those limits. For samples greater than 50 pairs, the formula might be converted from binomial to normal or beta functions and maximized.

Calculation of binomial coefficients when N is greater than about 15 becomes arduous. A table of coefficients (as their logarithms) for all values of N up to 100 may be found in [7].

Of more general interest, the results given here call attention to the danger of relying entirely on Weinberg's difference method even if the problem of correlated sex in DZ pairs [6] can be neglected or compensated. The Weinberg estimate magnifies the sampling variance of sex-concordance fourfold [equations (4)] and is thus much more variable than actual zygosity proportions. Hence the proportion of MZ twins in the source population, if accurately known, is more relevant for a small sample than is the Weinberg estimate. The method proposed here employs both sources of information.

Acknowledgment. The authors thank Dr. Karen D. Pettigrew for help in formulating this problem.

REFERENCES

1. Allen G (1981): Errors of Weinberg's difference method. In Gedda L, Parisi P, Nance WE (eds): Twin Research 3: Part A, Twin Biology and Multiple Pregnancy. New York: Alan R. Liss, pp. 71-74.
2. Allen G, Pettigrew KD (1973): Heritability of IQ by social class: evidence inconclusive. Science 182:1042-1044.
3. Bulmer MG (1970): The Biology of Twinning in Man. Oxford: Clarendon Press.
4. Eaves LJ, Jinks JL (1972): Insignificance of evidence for differences in heritability of IQ between races and social classes. Nature 240:84-88.
5. Heuser RL (1967): Multiple Births: United States 1964. Washington: US Government Printing Office (Public Health Service Publication No 1000, Series 21, No 14).
6. James WH (1979): Is Weinberg's differential rule valid? Acta Genet Med Gemellol 28:69-71.
7. Lentner C (ed) (1981): Geigy Scientific Tables. Basle: Ciba-Geigy.
8. Weinberg W (1901-1902): Beiträge zur Physiologie und Pathologie der Mehrlingsgeburten beim Menschen. Arch Ges Physiol 88:346-430.

Correspondence: Dr. Gordon Allen, 9326 West Parkhill Drive, Bethesda, MD 20814, USA.

https://doi.org/10.1017/S0001566000004682 Published online by Cambridge University Press