Department of Medical Genetics, New York State Psychiatric Institute
New York 32

COMMENTS ON THE ANALYSIS OF TWIN SAMPLES [1]

by
Gordon Allen

I. THE PARENT POPULATION

Many of the conclusions drawn from twin studies depend upon observed differences or assumed comparability between two twin samples (e. g., monozygotic and dizygotic), between a twin sample and the general twin population, or between the twin population and the total population. Much attention has been given to possible biological or cultural factors affecting these comparisons, but several relatively simple problems of sampling and enumeration still seem to deserve further study.

Frequency of Twins in the General Population

Among all reported births in the United States between 1928 and 1949, 1.108 per cent of pregnancies reaching the viable stage terminated in twin deliveries (Guttmacher, 1953). These twin pregnancies producied 2.19 per cent of all the babies. Strandskov and Edelen (1946) and others have shown that almost exactly a third of all pairs are of opposite sex. Data of this type are important for any study on large numbers of twins for two reasons: they afford an estimate of the number of twins one may hope to obtain from a given population, and they provide a check on the representativeness of the collected twins after the sample is complete. Because twins have a higher mortality rate than single-born persons, the frequency of twins at birth is an over-estimate of the frequency in a population of adults. Also, because identical twins have a higher mortality rate than fraternal twins, the proportion of fraternal pairs and, consequently, the proportion of opposite-sex pairs among newborn twins, will be exceeded among twins in later age groups. Some data are available on twin frequencies at later ages; Wilson and Jones (1931) found 1.4 per cent of twins among school children in four California cities, and Essen-Möller (1941) estimated that the frequency of twins in the total population is about 4/5 that

[1] This report is part of a study undertaken jointly by the Department of Medical Genetics, New York State Psychiatric Institute, and the Laboratory of Socio-environmental Studies, National Institute of Mental Health, Public Health Service, United States Department of Health, Education and Welfare. The author is on the staff of the National Institute of Mental Health.

Ricevuto il 31. XII. 1954.

at birth. More precise and up-to-date estimates are now possible, and estimates which relate twin frequency to sample age.

Twin frequencies are ordinarily expressed as a ratio of twins to single-born or of twins to all persons, and this ratio can change materially only in age periods when total mortality is high and twin mortality still higher. Data to be found in *Vital Statistics of the U. S.* (National Office of Vital Statistics, 1950) reveal that, in 1950, stillbirths together with deaths in the first four weeks of life, i.e., perinatal deaths, accounted for 82 per cent of all deaths in the first year. In a cohort subject to mortality rates prevailing at the various age levels in 1950, perinatal mortality would account for 59 per cent of all deaths in the first 15 years of life. It is reasonable to assume that in twins, also, the mortality after the first month is relatively low.

Furthermore, most of the excess mortality in twins is probably due to factors in gestation and birth. It may be assumed, therefore, that the difference between twins and nontwins becomes less important with increasing age. Some support for this assumption is found in data obtained through the courtesy of the National Office of Vital Statistics. For deaths in the first four weeks of life during January, February and March of 1950, 85.8 per cent of the single-born mortality occurred in the first week and 14.2 per cent in the last three weeks. Of the twin mortality, 90.4 per cent occurred in the first week and only 9.6 per cent in the remaining three weeks. The standard error of this difference between twins and nontwins is 0.73 per cent. While the mortality is 5.7 times as great in twins during the first week, it is 3.9 times as great in the next three weeks and, according to Record *et al.* (1952), only 2.3 times as great in the remainder of the first year. Even the last ratio, 2.3, represents mainly the second to sixth months, in which most of the remaining mortality occurs. In view of this prompt falling off of differential mortality, and the relatively small number of deaths after the neonatal period, it is probable that the proportion of twins in the population changes little after the age of one month, and is practically constant after the first year.

It is not difficult to correct twin birth frequencies for perinatal deaths in a population for which mortality is known approximately. According to 1950 data supplied by the National Office of Vital Statistics, combined stillbirths and neonatal deaths were 14.7 per cent for twins and 3.9 per cent for all births. When these mortality figures are applied to the above frequency of twins at birth, a figure of 1.94 per cent is obtained for the frequency among babies alive at one month. The British data of Record *et al.*, on infants followed for one year, agree closely with the U.S. figures during the first month, and may therefore be used to estimate changes in the remainder of the first year. This adjustment leaves a final figure of 1.90 per cent of twins after the first year. This fall from the twin rate at birth is large enough to alter at the five per cent level of significance the number of twins expected in a population of about 10,000; i.e., in a sample containing about 200 twins.

In about 11.5 per cent of twin pairs followed from birth through the neonatal period by Karn (1953), neither twin survived. At this rate, only 88.5 per cent of twin births would be represented by at least one living twin at the age of one month, or 1.01 pairs per 100 living infants at that age. In about 76 per cent of Karn's cases, both twins survived to one month, which would leave only 0.87 complete twin pairs per 100 living infants. Figures for the U. S. given by Yerushalmy and Sheerar (1940) agree well with Karn's, and indicate that these estimates should be widely representative.

The reporting of stillbirths, both twin and single, is known to be unreliable. For the purpose of estimating the twin population, therefore, it would be better when possible to start with the frequency of twin confinements resulting in at least one live birth, and to express this as a proportion of all confinements leading to live births. The correction required for differential mortality of twins would then be smaller and more precise.

Distribution of Zygosity Types in the Twin Population

Agreement between a twin sample and the parent population with respect to the proportion of opposite sex pairs or, when known, the proportion of monozygotic pairs, is perhaps the most useful evidence of unbiased sampling in a twin study. Mainly because monozygotic (MZ) twins are believed to have a higher mortality rate than dizygotic (DZ) twins (Yerushalmy and Sheerar, 1940; Strandskov and Askins, 1953), there has been a tendency to distrust the proportions found at birth in evaluating samples of adult twins. Even independent estimation of the proportions in such samples has been questioned, because some regard the differential method of Weinberg (1901) as too crude, especially when it is used without modification of the 1:1 sex ratio or on populations past middle life (Essen-Möller, 1941; Waterhouse, 1953). The following statistics reveal that, in the two above respects, uncorrected estimates should adequately describe the twin population for the purpose of comparison with a sample.

According to the data of Yerushalmy and Sheerar, the early mortality of same-sex twins is about 22 per cent greater than that of opposite-sex twins, which raises the proportion of opposite-sex pairs from 33.25 per cent at birth (Strandskov and Edelen, 1946) to 34.26 per cent at the age of one month. This difference would become significant at the five per cent level of confidence only in samples of over 8000 twin pairs. Even when this observed excess mortality of the same-sex group is entirely ascribed to the monozygotic pairs, the approximate survival rates are found to be 86 per cent for DZ twins and 80 per cent for MZ twins. The resulting proportion of MZ twins, 31.4 per cent, is still close enough to one-third that the latter figure can safely be used in connection with samples of up to 2000 twin pairs. In any study where these effects of mortality become significant, the proportions can be estimated satisfactorily by equating the DZ twins to twice the number of opposite-sex twins, in what may be called the simplified differential method of Weinberg.

A reasonable question arises, however, as to the applicability of Weinberg's differential method in older age groups. Mortality rates are higher in males, and any deviation from the 1:1 sex ratio would theoretically increase the proportion of same-sex pairs among the DZ twins above 50 per cent. If based on a 1:1 sex ratio, the differential method would then underestimate the number of same-sex DZ twins, and overestimate the number of MZ twins. As originally proposed by Weinberg, however, the calculation made use of the actual population sex ratio, when this was known, to obtain a corrected estimate for the number of same-sex, DZ pairs. As Weinberg suspected, and as the following calculationswill show, even this is a needless refinement.

According to the U.S. Census for 1950, the sex ratio remains above 95 males per 100 females up to age 65. For persons over 64 the sex ratio is 89.6 and for persons over 74 it is 82.6. When proportions of same-sex and opposite-sex DZ pairs are calculated on the assumption of a primary sex ratio of 100, one finds that even when mortality reduces the sex ratio to 80 males per 100 females (or 44.4 per cent males), 49.4 per cent of unbroken pairs should be of opposite sex. The resulting estimate of the proportion of MZ pairs is only 0.81 per cent too high, an error which becomes statistically significant only in samples of over 10,000 pairs of twins. Furthermore, if the sample includes individuals from broken and unbroken pairs in the same ratio as in the population, the proportion of cases from opposite-sex pairs is found to have undergone no change at all. This is illustrated in the table.

Table 1 - Effect of differential mortality of the sexes upon distribution of sex-concordance in aged twin pairs. Illustrated for a hypothetical population of 40,000 dizygotic twin pairs. Equal numbers of males and females are assumed at the start, and a sex ratio of 80 males per 100 females is assumed after 82 percent mortality

	Same-sex Male	Same-sex Female	Same-sex Total	Opposite-Sex
Pairs at Conception	10,000	10,000	20,000	20,000
Male Individuals	20,000		20,000	20,000
Female Individuals		20,000	20,000	20,000
After Mortality of 80% of all females, 84% of all males:				
Unbroken Pairs	256	400	656	640
Broken Pairs, Male Surviving	2,688		2,688	2,560
Broken Pairs, Female Surviving		3,200	3,200	3,360
All Male Survivors	3,200		3,200	3,200
All Female Survivors		4,000	4,000	4,000
All Survivors	3,200	4,000	7,200	7,200

Dependence of Twin Frequency on Parent Population and on Sample Age

As nearly as possible, standard twin statistics used in the evaluation of samples should, of course, represent the populations from which the samples are drawn. The twin rate of 2.19 per cent obtained in the first section on the basis of Guttmacher's figures, is an average for all U. S. births from 1928 to 1949. Although the correction for perinatal mortality to 1.94 per cent is based on 1950 data, the latter agree well with the data of Record *et al.* (1952) obtained in Birmingham, England, in 1947. Similary, twin mortality in 1936 in New York State exclusive of New York City (Yerushalmy and Sheerar) was only ten per cent higher than in the 1950 data. Therefore, the corrected twin frequency of 1.9 per cent after the first year should apply to the present United States population up to about age 25.

For older age groups and for samples not comparable to the total U. S. population, the twin frequency is probably different. Gedda (1951) has reviewed national and racial differences in twin rates at birth, and many of these differences are substantial; Strandskov (1945) has evaluated the difference in twin rates between colored and white populations in the U. S. In addition, the magnitude of the differential mortality of twins and nontwins is probably proportional to stillbirth and infant mortality rates, which differ widely between the present generation and older generations, from one geographical area to another, and probably among different socio-economic segments of the population. The increasing infant survival rate in the past generation may favor twins by reducing the numerical difference in mortality between twins and nontwins, but this has apparently been more than offset by the decline in twin frequency at birth. According to Guttmacher's analysis, this frequency averaged 1.16 per cent from 1928 to 1938, and 1.07 per cent from 1939 to 1949. This difference represents only an eleven-year interval, but is more than half as great as the entire correction for perinatal mortality derived in the first section.

More important is another effect which sample age may have on estimates of twin frequency. After the first year of life, mortality is not expected to alter the relative frequency of twins; in the U. S., this should remain about 1.9 per cent in the survivors of the cohorts born between 1928 and 1949. On the other hand, the proportion of complete pairs has been shown above to fall to 0.87 per 100 infants even by the age of one month. This proportion continues to decline as deaths separate more and more pairs. When a a cohort reaches the very oldest ages, there will be almost no complete pairs, and nearly all pairs will be represented by a single twin. If the mortality of twins after the first year of life is no greater than that of single-born persons, there will still be about 1.9 per cent twins. Therefore, the number of pairs represented will be nearly 1.9 pairs per 100 individuals instead of 1.01 as at the age of one month. The rate and degree of approach to the higher figure at successive ages will depend upon mortality rates and upon the average interval between the deaths of twin partners. Other factors to be considered are immigration and emigration. By separating partners without reducing the frequency of twin individuals, these factors will affect local populations in the same way as mortality, raising the frequency of pairs represented by at least one twin still closer to the frequency of twin individuals.

Thus, the number of twin pairs represented per 100 population is difficult to estimate in any but the earliest age groups; so, too, is the number of unbroken pairs. On the other hand, the frequency of twin individuals in the population is nearly constant at all ages after the first year of life, and can be rather accurately estimated in any population for which one knows the twin frequency at birth and the differential perinatal mortality. Consequently, twin frequencies are best expressed in terms of twin individuals, rather than in terms of twin pairs, and this principle applies, first of all, to the twin data used as standards of reference. In the analysis of data obtained from twin samples, adherence to the principle becomes important for additional reasons, which are to be discussed in the second part of this paper.

Summary, Part I

In estimating twin frequencies from birth and mortality data, it is necessary to make some statistical corrections, the extent of which is demonstrated for the general population of the United States. In this population, twins constitute 2.19 per cent of all babies born since 1928. This rate is reduced to about 1.9 per cent by excess twin mortality within the first year of life, while the 2:1 ratio of same-sex to opposite-sex pairs observed at birth remains virtually unchanged. Later mortality reduces the proportion of males, but not to such an extent — even at age 75 — as to necessitate modification of the 1:1 sex ratio in estimating the proportion of monozygotic pairs by Weinberg's differential method.

After the first year of life, there is no evidence for a significant difference between the mortality rates of twins and nontwins, so that the proportion of twin individuals in the population may be assumed to remain nearly the same at all ages. However, as pairs are broken by mortality and migration, the number of intact pairs is reduced at successive ages. Among people who survive to an advanced age, the relative frequency of pairs represented by at least one twin may be nearly twice as great as at birth. For comparative purposes, therefore, the most useful way to express twin data is in terms of individuals rather than in terms of twin pairs.

References, Part I

ESSEN-MOELLER, E. 1941. Tvillingforskning. *Svenska läk.-sällsk. forhandl.* 1-10. Cited by Gedda (1951), p. 166.

GEDDA, L. 1951. *Studio dei Gemelli.* Rome: Orizzonte Medico.

GUTTMACHER, A. F. 1953. The incidence of multiple births in man and some other unipara. *Obst. Gyn.* 2: 22-35.

KARN, M. N. 1953. Data of twin births occurring in two English metropolitan hospitals. *A.Ge.Me.Ge.* 2: 152-163.

RECORD, R. G., GIBSON, J. R., & MCKEOWN, T. 1952. Foetal and infant mortality in multiple pregnancy. *Jour. Obst. Gyn.* 59: 471-482.

STRANDSKOV, H. H., & ASKINS, J. E. 1953. An analysis of stillbirth frequencies among members of twin pregnancies within the total, the « white » and the « colored » populations of the United States. *A.Ge.Me.Ge.* 2: 164-180.

STRANDSKOV, H. H., & EDELEN, E. W. 1946. Monozygotic and dizygotic birth frequencies in the total, in the « white » and the « colored » U. S. populations. *Genetics* 31: 438-446.

WATERHOUSE, J.A.H. 1953. Twin Studies. In Sorsby's *Clinical Genetics.* St. Louis: C. V. Mosby & Co., pp. 35-46.

WEINBERG, W. 1901. Beiträge zur Physiologie und Pathologie der Mehrlingsgeburten beim Menschen. *Pflügers Arch.* 88: 346-430.

WILSON, P. T., & JONES, H. E. 1931. A study of like-sexed twins. I. The vital statistics and family data of the sample. *Human Biology* 3: 107-132.

YERUSHALMY, J., & SHEERAR, S. E. 1940. Studies on twins. II. On the early mortality of like-sexed and unlike-sexed twins. *Human Biology* 12:247-263.

SOMMARIO I

Nel valutare la frequenza gemellare dai dati di natalità e mortalità, è necessario effettuare alcune correzioni statistiche, la cui estensione viene dimostrata per la popolazione generale degli Stati Uniti.

In questa popolazione i gemelli costituiscono il 2,19% di tutti i bambini nati dal 1928. Questo rapporto è ridotto a circa 1,9% dall'eccesso di mortalità gemellare entro il primo anno di vita, mentre il rapporto di 2:1 fra le coppie dello stesso sesso e quello di sesso opposto osservato alla nascita rimane virtualmente immutato. La mortalità successiva riduce la proporzione di maschi, ma non in modo tale — anche a 75 anni di età — da richiedere una modificazione del rapporto dei sessi di 1:1 nel valutare la proporzione di coppie monozigotiche mediante il metodo differenziale di Weinberg.

Dopo il primo anno di vita, non vi sono prove di una differenza significativa fra gli indici di mortalità dei gemelli e non gemelli in modo che la proporzione di individui gemelli nella popolazione, si può ritenere che rimanga quasi costante a tutte le età. Tuttavia, in quanto delle coppie vengono rotte dalla mortalità e dalla migrazione, il numero di coppie intatte è ridotto nelle età successive. Fra le persone che sopravvivono fino ad un'età avanzata, la frequenza relativa di coppie rappresentate da almeno un gemello può essere quasi doppia di quella alla nascita. Per scopi comparativi perciò il modo più utile di esprimere i dati gemellari è in termini di individui piuttosto che in termini di coppie gemellari.

RÉSUMÉ I

Pour évaleur la fréquence gémellaire des données relatives à la natalité et à la mortalité, il est nécessaire d'effectuer quelques corrections statistiques, dont l'extension est démontrée par la population générale des Etats-Unis.

Dans cette population, les jumeaux constituent le 2,19% de tous les enfants nés depuis 1928. L'excès de mortalité gémellaire au cours de la première année d'existence, réduit ce pourcentage à environ 1,9%, tandis que le rapport 2:1 parmi les couples du même sexe et celles d'un sexe opposé, observé à la naissance, demeure virtuellement inchangé. La mortalité successive réduit la proportion des garçons, mais insuffisamment — même à l'âge de 75 ans — pour demander une modification du rapport des sexes de 1:1 dans l'évaluation de la proportion des couples MZ par la méthode différentielle de Weinberg.

Aucune preuve d'une différence significative n'existe après la première année d'existence, parmi les index de mortalité des jumeaux et des non jumeaux. Dès lors, tout permet de croire que la proportion de jumeaux demeure quasi constante à tous les âges. Toutefois, compte tenu que les couples sont séparés par la mortalité et par la migration, le nombre de couples intacts est réduit aux âges successifs. Parmi les personnes qui survivent jusqu'à un âge avancé, la fréquence relative des couples représentés par au moins un jumeau, peut être considérée quasi double de ce qu'elle n'est à la naissance. La méthode la plus utile pour établir des comparaisons est d'exprimer les données gémellaires en termes d'individus plutôt qu'en termes de couples gémellaires.

ZUSAMMENFASSUNG I

Wenn Geburten- und Sterbeziffern für die Berechnung der Zwillingshäufigkeit benützt werden, erfordern sie gewisse statistische Korrekturen, die für die amerikanische Gesamtbevölkerung dargelegt worden sind. Der Zwillingsprozentsatz in den Vereinigten Staaten beläuft sich seit 1928 auf 2,19% für alle Geburten. In Anbetracht der erhöhten Zwillingssterblichkeit im ersten Lebensjahre ist jedoch anzunehmen, dass sich dieser Prozentsatz auf ungefähr 1,9% verringert, während sich das ursprüngliche Zahlenverhältniss zwischen gleich- und ungleichgeschlechtigen Zwillingspaaren (2:1) kaum verändert. Wenn sich auch der Prozentsatz männlicher Zwillinge infolge höherer Mortalität im späteren Leben etwas erniedrigt, so ist diese Reduktion doch nicht bedeutend genug — nicht einmal in der Altersklasse von 75 Jahren — um in der Berechnung des Prozentsatzes eineiiger Zwillingspaare nach der Weinberg'schen Differenzierungsmethode eine Korrektur des ursprünglichen Geschlechtsverhältnisses (1:1) nötig zu machen.

Da die Sterblichkeit der Zwillinge vom zweiten Lebensjahre an kaum grösser ist als jene der Nicht-Zwillinge, so bleibt das Verhältnis der Zwillinge zur Allgemeinbevölkerung in allen Altersstufen annähernd das gleiche. Mit zunehmendem Alter werden jedoch Zwillinge durch Tod und Abwanderung immer häufiger getrennt, so dass sich die Zahl vollständiger Paare immer mehr verringert. In den höheren Altersklassen dürfte die Zahl einzelner Zwillinge ungefähr doppelt so gross sein wie zur Zeit der Geburt. Für vergleichende Zwillingsuntersuchungen ist es daher ratsam, mit Einzelzwillingen und nicht mit Zwillingspaaren zu rechnen.