



Received 20 April 1978
Final 28 August 1978

Testing Twin Means and Estimating Genetic Variance

Basic Methodology for the Analysis of Quantitative Twin Data

Joe C. Christian

Department of Medical Genetics, Indiana University School of Medicine, Indianapolis

Twin methodology for the analysis of continuously distributed traits is presented as a guideline for initial steps in the analysis of twin data. Tests for heterogeneity of twin means and variances are reviewed along with estimates of genetic variance.

Analysis of twin data appears on the surface to be a rather simple task, but investigators may well be confused by the wealth of choices available for hypothesis testing. This paper is designed to help the less experienced investigator, with perhaps a modest amount of data, begin analysis with maximum attention to testing all testable assumptions of the twin model, and using recent examples of applications of the methods.

Key words: Twin data analysis, Genetic variance, Associations with twin type, Twin placentation, Environmental covariance, Heritability estimates

Twins are commonly used to study continuously distributed or quantitative traits, but there is little agreement among investigators about how to analyze and present twin data. This paper is an attempt to answer the often asked question of where to start on twin analysis, but is far from inclusive of all possible analytic strategies.

No attempt is made to answer the question raised by responsible investigators as to whether there is any place for twins in studies of biometric traits in man [11, 19]. How-

This manuscript was reviewed by the members of the Working Group on Standardization of Twin Data Analysis of the International Society for Twin Studies, including (besides the present author) Drs. G. Allen (Rockville, Maryland, USA), G. Brenci (Rome, Italy), E. Defrise-Gussenhoven (Brussels, Belgium), R.C. Elston (Chapel Hill, North Carolina, USA) and R. Garrison (Bethesda, Maryland, USA). A review of multivariate methods of analysis, authored by D.W. Fulker (London, Great Britain), will follow.

This is publication no. 77-71 from the Department of Medical Genetics and was supported in part by the Indiana University Human Genetics Center, PHS P50 21054, and research grants PHS R01 HL 20034 and PHS P50 HL 14159.

ever, I have a firm conviction that twins will prove to be one of the most effective and cost-efficient methods of quantitative genetic analysis if attempts are made to prevent biases of ascertainment, to test all testable assumptions of the twin model, to use the most efficient methods of analysis available, and to finally test on singletons hypotheses formulated from twin studies.

TEST FOR ASSOCIATION OF A VARIABLE MEAN WITH TWIN TYPE

Comparisons of monozygotic (MZ) and dizygotic (DZ) twins are used to estimate genetic variance or heritability. Comparisons of monochorionic and dichorionic MZ twins are also made, as are comparisons of twin types divided by race, sex, etc. Whenever two types of twins are compared, an important early step in the analysis is a test of the differences between the means of the twin types. Discovery of a difference between twin means is evidence for an association between twin type and the trait being studied. Detection of such associations is important to prevent possible bias in further steps in the analysis and to provide insight into sources of variation for the trait being studied.

Christian and Norton [8] proposed that the appropriate test for significance of the difference between the means (\bar{Y}) of MZ and DZ twins (applicable to other types of twins) is the following t' test:

$$t' = (\bar{Y}_{MZ} - \bar{Y}_{DZ}) / \sqrt{(M_{AMZ}/2n_{MZ}) + (M_{ADZ}/2n_{DZ})} \tag{1}$$

with approximate degrees of freedom =

$$\frac{[(M_{AMZ}/2n_{MZ}) + (M_{ADZ}/2n_{DZ})]^2}{[(M_{AMZ}/2n_{MZ})^2/(n_{MZ} - 1)] + [(M_{ADZ}/2n_{DZ})^2/(n_{DZ} - 1)]} \tag{2}$$

This test is based upon the nested structure of twin data. Twin type is a fixed effect. The twin pairs nested within twin types constitute a random effect. Finally, individuals nested within twin pairs are another random effect. The appropriate test of the fixed effect (the difference between the means of the twin types) uses the mixed model nested or hierarchical analysis of variance [23: pp 266 ff].

The usual analysis of variance among and within twin pairs is shown in the Table and may be used as the first step in the nested analysis referred to above. The t' test (1) can then be calculated.

There is little data available about how common associations are between trait means and twin type. Many early studies used the independent-samples t test, making positive results suspect because this test tends to overestimate the significance of differences be-

TABLE. General Analysis of Variance Model for Twin Studies

Source of variation	Monozygotic pairs		Dizygotic pairs	
	df	Mean squares	df	Mean squares
Among pairs	$n_{MZ} - 1$	M_{AMZ}	$n_{DZ} - 1$	M_{ADZ}
Within pairs	n_{MZ}	M_{WMZ}	n_{DZ}	M_{WDZ}

df = Degrees of freedom; n_{MZ} = number of MZ pairs, n_{DZ} = number of DZ pairs.

tween twin means [8]. Havlik et al [15] used a test very similar to the t' test (an F test the equivalent of Christian and Norton's t_2 [8]) and found that 4 of 11 clinical chemistry values had evidence for unequal means of MZ and DZ twins ($P < 0.05$) in 514 pairs of adult male twins. These twins were studied in the US National Heart, Lung and Blood Institute (NHLBI) Collaborative Twin Study. The association of twin type and the means of quantitative traits may therefore be relatively common.

TEST FOR ASSOCIATION OF TOTAL VARIANCE WITH TWIN TYPE

Kemphorne and Osborne [16] proposed that the total variance of twin types be compared to search for associations between sources of variation and type of twin. Haseman and Elston [14] proposed the appropriate F' test:

$$F' = (M_{ADZ} + M_{WDZ}) / (M_{AMZ} + M_{WMZ}), \text{ or}$$

$$F' = (M_{AMZ} + M_{WMZ}) / (M_{ADZ} + M_{WDZ}) \quad (3)$$

with the larger sum in the numerator and the approximate degrees of freedom for the DZ sum calculated as follows:

$$\frac{(M_{ADZ} + M_{WDZ})^2}{[(M_{ADZ})^2/n_{DZ}^{-1}] + [(M_{WDZ})^2/n_{DZ}]} \quad (4)$$

and similarly for the MZ sum.

As with associations of twin type and the mean, associations of the variance of a trait and twin type may indicate sources of variation that seriously bias further analyses, such as estimating genetic variance or heritability. In the NHLBI Twin Study, 15 of 31 quantitative variables were found to have significant evidence ($P < 0.05$) for unequal total variance of MZ and DZ twins [9]. Reed et al [20] reported that 20 of 71 dermatoglyphic variables had significantly ($P < 0.05$) unequal total variances for MZ and DZ twins. Association of twin type with the variance of quantitative traits, therefore, does not appear to be unusual.

Possible causes for association of total variance with twin type are prenatal environmental influences due to placental types of MZ twins. After finding evidence for a smaller plasma cholesterol total variance for MZ compared to DZ twins in adults [5] and newborn infants [4], the newborn MZ twins were then divided into monozygotic and dizygotic pairs. The monozygotic pairs were found to have a within-pair mean square one-fifth of that of dizygotic MZ twins ($P < 0.05$ [10]). A similar finding was reported for plasma cholesterol variation in adult monozygotic and dizygotic MZ twins [6]. Evidence for unequal variance of monozygotic and dizygotic MZ twins has also been found for hematocrit [24], IgG [1], IQ [17], and dermatoglyphics [21, 22].

FUNCTIONS OF POPULATION GENETIC VARIANCE

The commonly used estimates of genetic variance from twin data (5 and 7) are a complex fraction of population genetic variance [2]. An estimate of population genetic variance may be obtained from the following estimates by assuming that dominance and epistatic deviations are zero and multiplying the twin estimates by two.

Several estimates of genetic variance have been proposed for use with twin data [2, 12, 14, 16]. For studies involving large numbers of twins, and for traits where there is no inequality of the total variances of MZ and DZ twins, all of these estimates are very similar. The within-pair estimate (\hat{G}_{WT}) is the most commonly used of these estimates:

$$\hat{G}_{WT} = M_{WDZ} - M_{WMZ} \quad (5)$$

The significance of \hat{G}_{WT} is tested by the following F ratio:

$$F = M_{WDZ}/M_{WMZ} \quad (6)$$

with the appropriate degrees of freedom for these two mean squares as shown in the Table.

If the trait being studied has evidence for association of its mean or variance with a twin type, then it is possible that \hat{G}_{WT} is a biased estimator. For some traits it may be possible to remove apparent associations of the variance and the mean with twin type by appropriate transformations.

Falconer [12] published an among-twin-pair component estimate of genetic variance, and Haseman and Elston [14] presented an unweighted least-squares estimate of genetic variance equal to two times the among-pair component estimate of Falconer. Christian et al [2] proposed that Falconer's estimate, designated \hat{G}_{CT} , is an unbiased estimator of genetic variance when the total variances of MZ and DZ twins are unequal due to environmental influences. \hat{G}_{CT} is calculated as follows:

$$\hat{G}_{CT} = [(M_{AMZ} - M_{ADZ}) + (M_{WDZ} - M_{WMZ})]/2 \quad (7)$$

with the following significance test:

$$F' = (M_{WDZ} + M_{AMZ})/(M_{ADZ} + M_{WMZ}) \quad (8)$$

$$\text{numerator df} = \frac{(M_{WDZ} + M_{AMZ})^2}{[(M_{WDZ})^2/n_{DZ}] + [(M_{AMZ})^2/n_{MZ}^{-1}]}$$

and similarly for denominator degrees of freedom.

Kempthorne and Osborne [16] reviewed several theoretical environmental influences that could cause the total variances of MZ and DZ twins to be unequal. The recent findings of placental type associated with total variance of quantitative traits is evidence for unequal environmental influences on twin types and a situation where \hat{G}_{CT} would be expected to give an unbiased estimate of genetic variance. Feinleib et al [13] further compared the properties of \hat{G}_{CT} and \hat{G}_{WT} when the environmental variances and covariances of MZ and DZ twins differ.

Nance, in a letter to the Editor of the American Journal of Human Genetics [18], gave an example of unequal total variances due to unequal genetic backgrounds of MZ and DZ twins and proposed use of \hat{G}_{WT} even though the variances of MZ and DZ twins are found to be unequal. It was subsequently pointed out [7] that this example would be expected to cause an association of twin type with twin means as well as variances.

There has been no estimate of genetic variance proposed to be unbiased when the means of MZ and DZ twins are different.

TEST OF EQUALITY OF ENVIRONMENTAL COVARIANCES OF MZ AND DZ TWINS

A serious possibility for bias of all commonly used estimates of genetic variance is inequality of the environmental covariances of MZ and DZ twins. If, for example, the environmental covariance of MZ twins is greater than the environmental covariance of DZ twins, then most estimates of genetic variance and heritability will be biased upward. To guard against spurious estimates of genetic variance, a screening F test of M_{ADZ}/M_{WDZ} has been suggested [3]. If an estimate of genetic variance is significant and this F ratio is not appreciably greater than 1.0, then the possibility of greater environmental covariance for MZ twins should be suspected.

HERITABILITY ESTIMATES FROM TWIN DATA

Numerous heritability estimates have been proposed for use with twin data. Heritability is a decimal fraction between 0 and 1.0 obtained by dividing an estimate of population genetic variance by an estimate of population variance. The estimates of genetic variance discussed in the previous section (\hat{G}_{WT} and \hat{G}_{CT}) are a complex fraction of population genetic variance, and only by assuming no dominance or epistasis can they be transformed into an estimate of population genetic variance by doubling them. It is also difficult to obtain an unbiased and stable estimate of population total variance, further complicating estimation of heritability.

REFERENCES

1. Byron E, Slavin B (1974): Serum IgG levels in fetio-fetal transfusion syndrome. *Arch Dis Childhood* 49:908–910.
2. Christian JC, Kang KW, Norton JA Jr (1974): Choice of an estimate of genetic variance from twin data. *Am J Hum Genet* 26:154–161.
3. Christian JC, Feinleib M, Norton JA Jr (1975): Statistical analysis of genetic variance in twins. *Am J Hum Genet* 27:807.
4. Christian JC, Kang KW, Nance WE (1975): Plasma cholesterol in newborn human twins (abstr). *J Am Oil Chem* 52:123a.
5. Christian JC, Feinleib M, Hulley SB, Castelli RR, Fabsitz RR, Garrison RJ, Borhani NO, Rosenmann RH, Wagner J (1976): Genetics of plasma cholesterol and triglycerides: A study of adult male twins. *Acta Genet Med Gemellol* 25:145–149.
6. Christian JC, Uchida IA (1976): Plasma cholesterol variation in monozygotic and dizygotic monozygotic twins (abstr). *Excerpta Medica Int Cong Ser* 397:28.
7. Christian JC, Kang KW, Norton JA Jr (1977): Comparison of within-pair and among-component estimates of genetic variance from twin data. *Am J Hum Genet* 29:208–210.
8. Christian JC, Norton JA Jr (1977): A proposed test of the difference between the means of monozygotic and dizygotic twins. *Acta Genet Med Gemellol* 26:49–54.
9. Christian JC (1978): The use of twin registers in the study of birth defects. In Summitt RL, Bergsma D (eds): "Cell Surface Factors, Immune Deficiencies, Twin Studies." New York: Alan R. Liss for The National Foundation—March of Dimes. *Birth Defects: Original Article Series XIV: 6A*, pp 167–178.
10. Corey LA, Kang KW, Christian JC, Norton JA Jr, Harris RE, Nance WE (1976): Effects of chorion type on variation in cord blood cholesterol of monozygotic twins. *Am J Hum Genet* 28:433–441.
11. Elston RC, Boklage CE (1978): An examination of fundamental assumptions of the twin method. In Nance WE, Allen G, Parisi P (eds): "Twin Research: Part A, Psychology and Methodology." New York: Alan R. Liss, pp 189–199.

12. Falconer DS (1960): "Introduction to Quantitative Genetics." New York: Ronald Press.
13. Feinleib M, Garrison RJ, Fabsitz R, Christian JC, Hrubec Z, Borhani NO, Kannel WB, Rosenman R, Schwartz JT, Wagner JO (1977): The NHLBI twin study of cardiovascular disease risk factors: Methodology and summary of results. *Am J Epidemiol* 106:284–295.
14. Haseman JK, Elston RC (1970): The estimation of genetic variance from twin data. *Behav Genet* 1:11–19.
15. Havlik R, Garrison R, Fabsitz R, Feinleib M (1977): Genetic variability of clinical chemical values. *Clin Chem (Winston-Salem, NC)* 23:659–662.
16. Kempthorne O, Osborne RH (1961): The interpretation of twin data. *Am J Hum Genet* 13:320–339.
17. Melnick M, Myrianthopoulos NC, Christian JC (in press): The effects of chorion type on variation in IQ in the NCPP twin population. *Am J Hum Genet*.
18. Nance WE (1976): Note on the analysis of twin data. *Am J Hum Genet* 28:297–299.
19. Rao DC, Morton NE, Yee S (1976): Resolution of cultural and biological inheritance by path analysis. *Am J Hum Genet* 28:228–242.
20. Reed T, Sprague FR, Kang KW, Nance WE, Christian JC (1975): Genetic analysis of dermatoglyphic patterns in twins. *Hum Hered* 25:263–275.
21. Reed T, Uchida IA, Christian JC (1976): Dermatoglyphic variation in monozygotic twins of differing placental types (abstr). *Excerpta Medica Int Cong Ser* 397:37.
22. Reed T, Uchida IA, Norton JA Jr, Christian JC (in press): Comparisons of dermatoglyphic patterns in monozygotic and dizygotic twins. *Am J Hum Genet*.
23. Sokal RR, Rohlf FJ (1969): "Biometry." San Francisco: WH Freeman.
24. Strong SJ, Corney G (1967): The placenta in twin pregnancy. Oxford: Pergamon Press.

Correspondence: Dr. J. C. Christian, Department of Medical Genetics, Indiana University School of Medicine, 1100 West Michigan Street, Indianapolis, IN 46202, USA.