GENETIC MODELS FOR THE ANALYSIS OF DATA FROM THE FAMILIES OF IDENTICAL TWINS

WALTER E. NANCE AND LINDA A. COREY[1]

Department of Human Genetics, Medical College of Virginia, Richmond, Virginia 23298[2]
and
Indiana University School of Medicine, Indianapolis, Indiana

Manuscript received September 5, 1975

ABSTRACT

Genetic models are described which exploit the unique relationships that exist within the families of identical twins to obtain weighted least squares estimates of additive, dominance and epistatic components of genetic variance as well as estimates of the contributions of *X*-linked genes, maternal effects and three sources of environmental variation. Since all of the relationships required to achieve a resolution of these variance components are contained within each family unit, the model would appear to be superior to previous approaches to the analysis of quantitative traits in man.

THE families of identical twins contain individuals who share all, one-half, one-quarter, and none of their genes. The presence of these relationships in every family group provides an unusual opportunity to detect and measure many important sources of human variation (NANCE *et al.* 1974, NANCE 1976). In this paper we describe a mathematical model for the analysis of quantitative data from these families. Numerical examples will be given in subsequent publications.

DESCRIPTION OF GENETIC MODELS

General Model

The general structure of the data is shown in Figure 1. Monozygotic twins possess identical sets of nuclear genes and, consequently, their children are related to each other in the same way as conventional half-sibs. The two sibships within each half-sibship often differ in size, mean age, and sex composition but because they are ascertained through twin parents who are necessarily the same age and sex, the expected size and average age of the sibships is the same, in contrast to conventional half-sibships resulting from illegitimacy, polygamy, death of a parent, or divorce. After the observations on the children have been adjusted for age, sex, and parity effects if necessary, analysis of variance permits

This is paper No. 1 from the Department of Human Genetics of the Medical College of Virginia. This work was initiated while the senior author was Principal Investigator of the Indiana University Human Genetics Center, GM 21054.

[1] NIH Postdoctoral Fellow PSH DE 00007, Department of Medical Genetics, Indiana University School of Medicine.

[2] Present address of both authors.

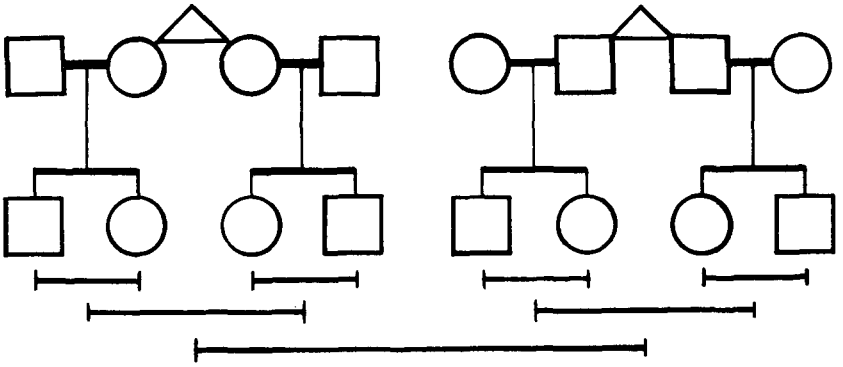

FIGURE 1.—Diagram showing structure of data. Lines below pedigree suggest how observed variation in offspring may be partitioned by a nested analysis of variance into among-half-sibship, between-sibship—within-half-sibship and within-sibship mean squares.

partitioning of the total sum of squares into that accounted for by half-sibships, by full-sibships nested within half-sibships, and by individuals within full-sibships. Since the number of children within each sibship is variable, numerical approximations for the coefficients (b_1, b_2, and a) of the between-sibship within-half-sibship (σ_B^2), and among-half-sibship (σ_A^2) components of variation must be calculated from the observed distribution of family size as shown in Table 1 (SNEDECOR 1956). Once the coefficients are determined, the variance components can be estimated from the mean squares in the usual manner. The genetic interpretation of the variance components of the half-sib analysis of variance is shown in Table 2. The model follows the conventional half-sib analysis of variance (FALCONER 1961; COMSTOCK 1955), except that the availability of both maternal and paternal half-sibships permits the inclusion of maternal effects in the model. Since the genetic expectations for the within-sibship mean squares obtained from the analyses of variance of families of male and female monozygotic twins are identical, these mean squares can be tested for heterogeneity and

TABLE 1

Analysis of variance of data from offspring of identical twins

Source of variation	D.F.	Mean square	Expected mean square
Among half-sibships	N—1	*MSA*	$\sigma^2_W + b_2\sigma^2_B + a\sigma^2_A$
Between sibships—within half-sibships	N	*MSB*	$\sigma^2_W + b_1\sigma^2_B$
Within sibships	$\Sigma(n_{i1} + n_{i2})-2n$	*MSW*	σ^2_W
Total	$\Sigma(n_{i1} + n_{i2})-1$	*MST*	

$$a = \frac{\Sigma(n_{i1}+n_{i2}) - \Sigma(n_{i1}+n_{i2})^2/\Sigma(n_{i1}+n_{i2})}{N-1} \qquad b_1 = \frac{\Sigma(n_{i1}+n_{i2}) - \Sigma(n_{i1}^2+n_{i2})^2/\Sigma(n_{i1}+n_{i2})}{N}$$

$$b_2 = \frac{\Sigma(n_{i1}^2+n_{i2})^2/(n_{i1}+n_{i2}) - \Sigma(n_{i1}^2+n_{i2})^2/\Sigma(n_{i1}+n_{i2})}{N-1}$$

TABLE 2

Genetic interpretation of variance components from offspring analyses of variance: General model

Variance component	Covariance estimated	Genetic interpretation								
		V_A	V_D	V_{AA}	V_{AD}	V_{DD}	V_M	V_{EH}	V_{ES}	V_{EW}
[1] $\sigma^2{}_A$ ♂	covHS ♂	1/4	0	1/16	0	0	0	1	0	0
[2] $\sigma^2{}_B$ ♂	covS ♂ —covHS ♂	1/4	1/4	3/16	1/8	1/16	1	0	1	0
[3] $\sigma^2{}_W$	V_T—covS	1/2	3/4	3/4	7/8	15/16	0	0	0	1
[4] $\sigma^2{}_B$ ♀	covS ♀ —covHS ♀	1/4	1/4	3/16	1/8	1/16	0	0	1	0
[5] $\sigma^2{}_A$ ♀	covHS ♀	1/4	0	1/16	0	0	1	1	0	0

pooled if homogeneous, leading to a total of five equations from the two classes of kinships.

On the average, half-sibs share one-quarter of their genes, and since all of the common elements are derived from a single parent, or in the present case from genetically identical parents, the among half-sibship component of variance, $\sigma^2{}_A$ (which is equivalent to the covariance of half-sibs), contains ¼ of the additive genetic variance (V_A) associated with the trait. The half-sib covariance also includes 1/16 of the variance (V_{AA}) resulting from the epistatic effects of additive genes (COMSTOCK 1955). Finally, although it may be possible to assume there is no environmental variation among half-sibships in experimental animals (MATHER and JINKS 1971), this would clearly be an oversimplification for many human traits. For this reason, an among-half-sibship environmental component, V_{EH}, is also included in the model (Table 2).

The within-sibship mean square, $\sigma^2{}_W$, is equivalent to the total phenotypic variance minus the covariance of full sibs. Therefore, it includes $1/2V_A$ and $3/4V_D$ in addition to the epistatic interaction components shown in Table 2. The within-sibship environmental component, V_{EW}, may be thought of as a measure of environmental variation that affects all individuals irrespective of their genetic relationship. This component includes the experimental error variance as well as variation resulting from age, sex, or parity effects, unless these are removed prior to analysis. If present, genetic-environmental interactions could inflate the environmental variance (V_{EW}), while the existence of a significant covariance between genotype and environment would tend to bias the genetic components of variance.

The between-sibship within-half-sibship component, $\sigma^2{}_B$, is equivalent to the full-sib covariance minus the half-sib covariance. Consequently, it includes $1/4V_A$ and $1/4V_D$ as well as the epistatic interaction components given in Table 2. As with the two previous components, a source of environmental variation, V_{ES}, is included in the model to take into account possible environmental differences between-sibships within-half-sibships.

Maternal effects can be detected by contrasting the analysis of variance for monozygotic male and female twin half-sibships, respectively. To the extent that genetically determined pre- or postnatal maternal effects influence a trait, the

offspring of female monozygotic twins would be expected to resemble one another to a greater extent than the children of male twins. Consequently, in the female twin half-sibships, the maternal effect, V_M, appears in the among-half-sibship component, $\sigma^2{}_A$♀. In contrast, since the children of male twins are born to genetically unrelated mothers, if maternal effects exist, they will augment $\sigma^2{}_B$♂ in comparison to $\sigma^2{}_B$♀. Since all the offspring within a sibship are born to the same mother, genetically determined maternal variation does not enter into the within-sibship mean square, and as noted previously, the within-sibship sums of squares from the families of male and female twins may, therefore, be combined to obtain a pooled estimate of $\sigma^2{}_W$. In this model, the maternal component includes variation arising from maternal dominance effects (V_{MD}), from the covariance between direct additive and maternal effects (V_{AM}) and from the covariance between direct dominance and maternal effects (V_{DM}), in addition to first order maternal effects. Note that the five equations shown in Table 2 are not linearly independent, since the sum of the coefficients of the first two equations is identical to that of the last two. Consequently, no more than four parameters could be estimated from these five component equations alone.

At least four additional equations may be obtained from the parental data. As shown in Table 3, the total sum of squares from the twin parents may be partitioned by a conventional analysis of variance into an among- and a within-twin pair component, in which the within-pair component will estimate the within-pair environmental variance, while the among-pair component will include all of the genetic variance as well as the variance attributable to environmental differences among pairs. Biases in the estimation of variance components may also be introduced here by the presence of significant genetic-environmental interaction and/or covariance.

A covariance between husband and wife may result either because of environmental similarities or because of nonrandom mate selection. The direction of the effect of nonrandom mating will be positive or negative, depending upon whether mating is assortative or disassortative. However, in either case, nonrandom mating will lead to a positive covariance between spouses. Although the effects of nonrandom mating (V_{AS}) are not estimated directly in the subsequent analysis, the relationships shown in Table 3 permit a clear separation to be made

TABLE 3

Genetic interpretation of parental data: General model

Relationship	Parameter	Genetic interpretation									
		V_A	V_D	V_{AA}	V_{AD}	V_{DD}	V_M	V_{EH}	V_{ES}	V_{EW}	V_{AS}
[6] Among twin pairs	$\sigma^2{}_{AT}$	1	1	1	1	1	1	1	0	0	0
[7] Within twin pairs	$\sigma^2{}_{WT}$	0	0	0	0	0	0	0	1	1	0
[8] Husband-Wife covariance	covHW	0	0	0	0	0	0	1	1	0	1
[9] Spouse-Spouse covariance	covSS	0	0	0	0	0	0	1	0	0	$\frac{V_{AS}}{V_T}$

between the contributions of common environment and assortative mating to the husband-wife covariance.

An additional set of five equations can be derived from the two conventional and three unique parent-offspring relationships that exist within the data. For example, the phenotypic covariance between the child and the twin parent differs from that between the child and the twin aunt or uncle only because of the increased environmental similarity of the former; while the covariance between the offspring and the genetically unrelated aunt or uncle provides another estimate of V_{EH}, the among-half-sibship environmental component. It seems likely that for some traits the maternal effects which are detected in the offspring analysis may not be equivalent to those measured by the parent-offspring covariance. This difference is evident when the expectations for the maternal effects in the two relationships are compared. Based on KEMPTHORNE (1957), the genetic expectation for the among-female-half-sibship component of variance can be written as

$$1/4V_A + 1/16V_{AA} + V_{MA} + V_{AM} + V_{MD} ,$$

while the expectation for the mother-offspring covariance is

$$1/2V_A + 1/4V_{AA} + 1/2V_{MA} + 5/4V_{AM} + V_{DM} .$$

In the present model, the epistatic (V_{AM}, V_{DM}) and dominance maternal effects (V_{MD}) are pooled with the linear (additive maternal) effects (V_{MA}) in both relationships to obtain an overall maternal effect (V_M). For this reason, it would be desirable to search for heterogeneity in the estimates of V_M derived from the two sets of equations prior to conducting an analysis over all of the equations.

Data Analysis

An equation of estimation can be derived from each of the fourteen expectations described in Tables 2–4. Measurement of the genetic and environmental parameters of interest can then be achieved by the simultaneous solution of appropriate sets of the foregoing equations by the least squares method. Two approaches are possible, a weighted or an unweighted analysis. For the latter, variance components can be used, and each component is assumed to be statistically independent and known with equal accuracy. In practice, the weighted analysis is preferable since it permits hypothesis testing and the estimation of confidence intervals. However, mean squares must be used in the equations of estimation rather than variance components when this approach is followed. The equations are then weighted by the reciprocal of their sampling variances and constraints are applied to adjust for non-independence of the mean squares prior to obtaining the least squares solutions.

The equations of estimation may be represented by

$$\mathbf{M}\,\mathbf{G} = \mathbf{C}, \qquad (1)$$

where **M** is the m by n matrix of variable coefficients for m equations in n un-

TABLE 4

Genetic interpretation of parent-offspring covariances: General model

Relationship	Parameter	Genetic interpretation				
		V_A	V_{AA}	V_M	V_{EH}	V_{ES}
[10] Offspring-Mother	covOM	1/2	1/4	1	1	1
[11] Offspring-Father	covOF	1/2	1/4	0	1	1
[12] Offspring-Spouse	covOS	0	0	0	1	0
[13] Offspring-Twin Aunt	covOA	1/2	1/4	1	1	0
[14] Offspring-Twin Uncle	covOU	1/2	1/4	0	1	0

knowns, **G** is the n by one column vector of genetic and environmental parameters whose estimates are desired and **C** is the m by one column vector of variances and covariances derived from the data. When mean squares are used instead of variance components, a different coefficient matrix, **M**, must be calculated for each set of families because of unequal sibship sizes. This is accomplished by multiplying the coefficients of the components given in Tables 2–4 by the appropriate weights and accumulating like terms, a process which may be represented algebraically by

$$\mathbf{M} = \mathbf{A}\,\mathbf{K}, \tag{2}$$

where **K** is the original matrix of component coefficients of the 14 equations specified in Tables 2–4,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_7 \end{pmatrix},$$

$$\mathbf{A}_1 = \begin{pmatrix} a\hat{\sigma} & b_2\hat{\sigma} & 1 & 0 & 0 \\ 0 & b_1\hat{\sigma} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & b_1\hat{\varphi} & 0 \\ 0 & 0 & 1 & b_2\hat{\varphi} & a\hat{\varphi} \end{pmatrix}, \text{ and } \mathbf{A}_2 = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$$

One effect of adjusting the coefficients for unequal family size is that the rank of the first five equations is almost invariably increased from four to five and the rank of the first seven equations from five to seven.

The next step is to calculate the m by m **V** matrix of weights and constraints. **V** is a symmetric matrix whose diagonal elements are calculated from one of the following two formulae:

$$v_{ii} = \frac{2c_{i(k)}^2}{df + 2}, \tag{3(a)}$$

where $c_{i(k)}$ is the mean square whose variance v_{ii} is to be estimated, and df is the associated degrees of freedom, or

$$v_{ii} = \frac{c_{i(kl)}^2 + d_{(k)}d_{(l)}}{df + 2} , \qquad 3(b)$$

where $c_{i(kl)}$ is the mean cross product with df degrees of freedom whose variance is to be estimated and $d_{(k)}$ and $d_{(l)}$ are the sample variances of the two variables involved in the covariance (MATHER and JINKS 1971).

The off-diagonal elements of **V** are zero in the case of orthogonal contrasts or with completely independent data sets, as in the case of the first five equations. Otherwise, one of the following formulae are used:

For a covariance between two variances, $c_{i(k)}$ and $c_{j(l)}$, between a variance and an intraclass covariance, $c_{i(k)}$ and $c_{j(ll)}$, or between two intraclass covariances, $c_{i(kk)}$ and $c_{j(ll)}$:

$$v_{ij} = \frac{2d_{(kl)}^2}{df + 2} , \qquad 3(c)$$

where $d_{(kl)}$ is the sample covariance between the elements k and l.

For a covariance between a covariance and either a variance or an intraclass covariance,

1) with no elements in common, i.e., $c_{i(lm)}$ and $c_{j(k)}$ or $c_{i(lm)}$ and $c_{j(kk)}$,

$$v_{ij} = \frac{2d_{(km)}\, d_{(kl)}}{df + 2} ; \qquad 3(d)$$

2) with one element in common, i.e., $c_{i(k)}$ and $c_{j(kl)}$,

$$v_{ij} = \frac{2c_{i(k)}\, d_{(kl)}}{df + 2} , \quad \text{or} \qquad 3(e)$$

$$v_{ij} = \frac{2d_{i(k)}\, d_{(kl)}}{df + 2} . \qquad 3(f)$$

For a covariance between two covariances,

1) with no elements in common, i.e., $c_{i(kl)}$ and $c_{j(mn)}$,

$$v_{ij} = \frac{c_{i(kl)}\, c_{j(mn)} + d_{(km)}\, d_{(ln)}}{df + 2} ; \qquad 3(g)$$

2) with one element in common, i.e., $c_{i(kl)}$ and $c_{j(km)}$,

$$v_{ij} = \frac{c_{i(kl)}\, c_{j(km)} + d_{(k)}\, d_{(lm)}}{df + 2} , \qquad 3(h)$$

$$v_{ij} = \frac{d_{(kl)}\, d_{(km)} + d_{(k)}\, d_{(lm)}}{df + 2} , \qquad 3(i)$$

$$v_{ij} = \frac{c_{i(kl)}\, d_{(km)} + d_{(k)}\, d_{(lm)}}{df + 2} , \quad \text{or} \qquad 3(j)$$

$$v_{ij} = \frac{c_{i(kl)}\, c_{j(km)} + d_{(k)}\, c_{(lm)}}{df + 2} \quad . \qquad 3(k)$$

In all cases, c_i is used to designate elements of the **C** vector which are recalculated during each step of the iteration process (see below) while $d_{(i)}$ and $d_{(ij)}$ refer to sample variances and covariances which are not elements of the **C** vector and are therefore held constant during the iteration procedure. The formulae given for the off-diagonal elements are derived from those given by ELSTON (1975). An explicit definition of **V** is presented in Table 5.

The subsequent analysis is outlined below:

$$\mathbf{V}^{-1}\,\mathbf{M}\,\mathbf{G} = \mathbf{V}^{-1}\,\mathbf{C} \qquad (4)$$

$$\mathbf{M}'\,\mathbf{V}^{-1}\,\mathbf{M}\,\mathbf{G} = \mathbf{M}'\,\mathbf{V}^{-1}\,\mathbf{C}$$

$$(\mathbf{M}'\,\mathbf{V}^{-1}\,\mathbf{M})^{-1}\,\mathbf{M}'\,\mathbf{V}^{-1}\,\mathbf{M}\,\mathbf{G} = (\mathbf{M}'\,\mathbf{V}^{-1}\,\mathbf{M})^{-1}\,\mathbf{M}'\,\mathbf{V}^{-1}\,\mathbf{C}$$

$$\hat{\mathbf{G}} = (\mathbf{M}'\,\mathbf{V}^{-1}\,\mathbf{M})^{-1}\,\mathbf{M}'\,\mathbf{V}^{-1}\,\mathbf{C}$$

After each side of the equation has been weighted by the inverse of **V**, each side is premultiplied by the transpose of **M**; the resulting product is inverted and both sides of the equation are premultiplied by this inverse to obtain the least squares estimates $\hat{\mathbf{G}}$. The elements of **V** should be calculated not from the observed values of **C** but from their expected values which are based on the weighted estimates of $\hat{\mathbf{G}}$ (HAYMEN 1960). Consequently, the final values of $\hat{\mathbf{G}}$ must be calculated in an iterative manner: an initial estimate of **V** is made based on the observed values of **C**. From this, an initial solution for $\hat{\mathbf{G}}$ is obtained, from which the expected value of **C** is calculated. A new **V** matrix, based upon this value, is obtained, and

TABLE 5

Elements of the **V** *matrix: General model*

c_i	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]
[1] Among ♂	(a)	0	0	0	0	(c)	0	(d)	(c)	0	(e)	(e)	0	(e)
[2] Between ♂	0	(a)	0	0	0	0	(c)	(d)	0	(e)	0	0	0	0
[3] Within	0	0	(a)	0	0	0	0	0	0	0	0	0	0	0
[4] Between ♀	0	0	0	(a)	0	0	(c)	(d)	0	0	(e)	0	0	0
[5] Among ♀	0	0	0	0	(a)	(c)	0	(d)	(c)	(e)	0	(e)	(e)	0
[6] Among T	(c)	0	0	0	(c)	(a)	0	(e)	(c)	(f)	(f)	(d)	(f)	(f)
[7] Within T	0	(c)	0	(c)	0	0	(a)	(e)	0	0	0	0	0	0
[8] covHW	(d)	(d)	0	(d)	(d)	(e)	(e)	(b)	(h)	(h)	(h)	(h)	(j)	(j)
[9] covSS	(c)	0	0	0	(c)	(c)	0	(h)	(b)	(i)	(i)	(d)	(g)	(g)
[10] covMO	0	(e)	0	0	(e)	(f)	0	(h)	(i)	(b)	(h)	(h)	(j)	(j)
[11] covFO	(e)	0	0	(e)	0	(f)	0	(h)	(i)	(h)	(b)	(h)	(j)	(i)
[12] covSO	(e)	0	0	0	(e)	(d)	0	(h)	(d)	(h)	(h)	(b)	(k)	(k)
[13] covAO	0	0	0	0	(e)	(f)	0	(j)	(g)	(j)	(j)	(k)	(b)	0
[14] covUO	(e)	0	0	0	0	(f)	0	(j)	(g)	(j)	(j)	(k)	0	(b)

Letters refer to formulae given in Equations 3(a)-(k) in text.

the process is repeated until convergence is achieved. At the conclusion of the iteration, the variance-covariance matrix of $\hat{\mathbf{G}}$ may be calculated from the relationship:

$$\text{Var } \hat{\mathbf{G}} = (\mathbf{M}^1 \mathbf{V}^{-1} \mathbf{M})^{-1} . \quad (5)$$

The diagonal elements of this matrix are the variances of $\hat{\mathbf{G}}$ and may be used to establish approximate confidence intervals for the estimated parameters, based on large sample theory and the assumption that the original variables are normally distributed. In practice, the elements of **C** often differ by considerably more than an order of magnitude. As a consequence, it may be difficult to invert and iterate **V**, and the use of a large computer, double precision and a robust inversion routine are recommended.

Model for Sex-Linked Genes

The effects of *X*-linked genes may be detected by subdividing the kinships into half-fraternities and half-sororities. The genetic interpretation of the resulting variance components follows that given by MATHER and JINKS (1971) and is shown in Table 6. The effects of autosomal genes, maternal influence and environmental factors are distributed just as in the previous model and in an identical manner in sons and daughters. In hemizygous males, there can be no dominance effects of *X*-linked genes (V_{DX}), and the total variance resulting from the additive effects of *X*-linked genes (V_{AX}) is twice as great in males as in females, since the breeding value for *X*-linked traits in females is determined by the mean effect of a sample of two *X* chromosomes. Fathers make no contribution to the variation in *X*-linked additive effects in their sons, and consequently, the coefficient for the additive variance in the genetic expectation of the among-half-sibship component in the sons of male twins and the between-sibship within-half-

TABLE 6

Genetic interpretation of variance components from offspring analyses of variance: Model for X-linked genes

Variance component	Covariance estimated	Genetic interpretation							
		V_A	V_{AX}	V_D	V_{DX}	V_M	V_{EH}	V_{ES}	V_{EW}
Half fraternities									
[1] σ^2_A ♂	covHS ♂	1/4	0	0	0	0	1	0	0
[2] σ^2_B ♂	covS ♂ —covHS ♂	1/4	1	1/4	0	1	0	1	0
[3] σ^2_W	V_T—covS	1/2	1	3/4	0	0	0	0	1
[4] σ^2_B ♀	covS ♀ —covHS ♀	1/4	0	1/4	0	0	0	1	0
[5] σ^2_A ♀	covHS ♀	1/4	1	0	0	1	1	0	0
Half sororities									
[6] σ^2_A ♂	covHS ♂	1/4	1/2	0	0	0	1	0	0
[7] σ^2_B ♂	covS ♂ —covHS ♂	1/4	1/4	1/4	1/2	1	0	1	0
[8] σ^2_W	V_T—covS	1/2	1/4	3/4	1/2	0	0	0	1
[9] σ^2_B ♀	covS ♀ —covHS ♀	1/4	1/2	1/4	1/2	0	0	1	0
[10] σ^2_A ♀	covHS ♀	1/4	1/4	0	0	1	1	0	0

sibship component for the sons of female twins is zero. For daughters, the greatest proportion of the additive variance is expected in the among component for male twins and the between-within component for female twins. Finally, a greater proportion of the dominance effect is seen in the between-within component for daughters of both classes of twins than is found with autosomal genes.

The equations describing the relationships among the parents are closely similar to those given previously except that in the present model the among-twin-pair components for male and female twins have different expectations and, therefore, must be separated (Table 7). A total of nine distinct parent-offspring relationships can be extracted from the data. The genetic interpretation of these covariances is shown in Table 8. In all, 24 equations can be derived; however, the expectations for the spouse-spouse covariance (equation 15) and the offspring-spouse covariance (equation 20) are the same and should not be used simultaneously if full rank is to be maintained and matrix singularity avoided. In the **V** matrix for the offspring equations that is shown in Table 9, the off-diagonal elements relating mean squares should all be zero if each twin half-sibship is counted only as a half-fraternity or a half-sorority. However, if a given kinship contains boys and girls in both families, it may be used twice and the degrees of freedom in the off-diagonal element would refer to the number of such families. The remainder of the analysis proceeds as described previously.

DISCUSSION

Inferences about the genetic determination of quantitative traits in man are often drawn by comparing the correlation between relatives of various degree or, in the case of threshold traits, the incidence of the disease or trait among the relatives of an affected proband. In experimental organisms, where mating can be controlled and where environmental effects can either be manipulated by the investigator or effectively randomized, it is a relatively simple matter to measure the contributions of genetic and environmental effects. However, research on quantitative genetics in man is necessarily observational in nature, and in many of the conventional approaches, genetic and environmental effects are either inextricably confounded or inappropriately ignored by the method of analysis. For

TABLE 7

Genetic interpretation of parental data: Model for X-linked genes

Relationship	Parameter	Genetic interpretation								
		V_A	V_{AX}	V_D	V_{DX}	V_M	V_{EH}	V_{ES}	V_{EW}	V_{AS}
[11] Among ♂ twin pairs	$\sigma^2_{AT\,♂}$	1	1	1	1	1	1	0	0	0
[12] Among ♀ twin pairs	$\sigma^2_{AT\,♀}$	1	2	1	0	1	1	0	0	0
[13] Within twin pairs	σ^2_{WT}	0	0	0	0	0	0	1	1	0
[14] Husband-Wife covariance	covHW	0	0	0	0	0	1	1	0	1
[15] Spouse-Spouse covariance	covSS	0	0	0	0	0	1	0	0	$\frac{V_{AS}}{V_T}$

TABLE 8

Genetic interpretation of parent-offspring covariances: Model for X-linked genes

Relationship	Parameter	Genetic interpretation				
		V_A	V_{AX}	V_M	V_{EH}	V_{ES}
[16] Father-Son	covFS	1/2	0	0	1	1
[17] Twin uncle-Son	covUS	1/2	0	0	1	0
[18] Mother-Son	covMS	1/2	1	1	1	1
[19] Twin aunt-Son	covAS	1/2	1	1	1	0
[20] Offspring-Spouse	covOS	0	0	0	1	0
[21] Twin aunt-Daughter	covAD	1/2	1/2	1	1	0
[22] Mother-Daughter	covMD	1/2	1/2	1	1	1
[23] Twin uncle-Daughter	covUD	1/2	1	0	1	0
[24] Father-Daughter	covFD	1/2	1	0	1	1

TABLE 9

Elements of the **V** *matrix offspring equations from model for X-linked genes*

C_i	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Sons										
[1] Among ♂	(a)	0	0	0	0	(c)	0	0	0	0
[2] Between ♂	0	(a)	0	0	0	0	(c)	0	0	0
[3] Within	0	0	(a)	0	0	0	0	0	0	0
[4] Between ♀	0	0	0	(a)	0	0	0	0	(c)	0
[5] Among ♀	0	0	0	0	(a)	0	0	0	0	(c)
Daughters										
[6] Among ♂	(c)	0	0	0	0	(a)	0	0	0	0
[7] Between ♂	0	(c)	0	0	0	0	(a)	0	0	0
[8] Within	0	0	0	0	0	0	0	(a)	0	0
[9] Between ♀	0	0	0	(c)	0	0	0	0	(a)	0
[10] Among ♀	0	0	0	0	(c)	0	0	0	0	(a)

Letters refer to formulae given in Equations 3(a) and 3(c) in text.

example, in classical twin studies, the within-pair differences of monozygotic and dizygotic twins are compared under the assumption that the environmental similarities and differences of the two classes of twins are identical. Since the monozygotic and dizygotic twin pairs almost never live in the same homes, this assumption is demonstrably false in detail and implausible in the aggregate. It is perhaps not surprising that, when examined, the total variances of monozygotic and dizygotic twins are often found to be dissimilar, thus calling into question the applicability of the classical twin model. However, it is equally erroneous to assume that a difference in total variance necessarily implies the existence of significant environmental differences between the two classes of twins (CHRISTIAN, KANG and NORTON 1973), since genetic differences both among and within population groups are known to influence the occurrence of dizygotic twinning (BULMER 1970) and could equally well account for an inequality of the total variances. These differences are thought to arise, at least in part, from genetically

determined endocrinologic differences in some mothers of dizygotic twins (MILHAM 1964; BENIRSCHKE and KIM 1973). Whether these differences make a significant contribution to the total variation of dizygotic twins, either as a prenatal maternal effect or as a direct effect in later life, would obviously depend upon the trait in question. All too often, however, difficulties of this type are concealed by the use of correlation coefficients. As noted by WOLF (1950), the use of correlation coefficients can mask massive environmental effects, particularly in the analysis of parent-offspring data. In addition, the classical twin model does not permit a resolution of additive and dominance effects or the detection of epistasis and is sensitive to bias when genetic-environmental interactions and/or covariances are present. Because of these limitations, some authors have advocated the collection and analysis of data from rare or abnormal human relationships such as identical twins reared apart (JENSEN 1974), adopted children (HORN, LOEHLIN and WILLERMAN 1974; MUNSINGER 1975) or conventional and illegitimate half-sibs (RAO, MORTON and YEE 1974). Data from these heterogeneous sources are sometimes pooled with observations on normal twins and families in the mistaken belief that only in this way can a resolution of genetic and environmental effects be achieved (JINKS and FULKER 1970). The question of the comparability of environmental effects in these groups is seldom raised and the relevance of inferences drawn from abnormal relationships to the general population is often not considered (TIZARD and REES 1974).

In contrast, the MZ twin half-sib model is based upon the analysis of data on the normal children of identical twins, who are raised in their own homes by their biologic parents in a manner which would appear to differ in no important way from the families of singletons in the general population. Although there are doubtless biases in any sample of cooperative families, other than being selected for fertility, there would seem to be few intrinsic biases in families ascertained through identical twin parents, since monozygotc twinning occurs with approximately equal frequency in all racial groups and is not known to be associated in a straightforward manner with maternal age, fertility drugs, or other environmental or genetic factors. In the present model, all of the genetic relationships that are needed to separate genetic and environmental effects and to partition the genetic variance into its additive, dominance, epistatic and maternal components are contained within each kinship. No assumptions need be made about the comparability of environmental effects in different classes of relatives, since the environmental effects acting on the parents may be partitioned into exactly the same components as in the offspring analysis. This is because the twin parents and their spouses generally live with their own children in different homes, and many potentially important sources of environmental variation such as diet, socioeconomic status, geographic location, climate, religion, or exposure to culture, books, pollution, or trace elements will tend to be distributed among and within kinships, with respect to the parents, in the same way that they are distributed for the children.

The ability to partition the environmental variance may prove to be an important feature of the model. The observed distribution of the environmental vari-

ance for a given trait could lead to the formulation and testing of specific hypotheses about the sources of environmental variation. The detection and estimation of genetic-environmental interactions is an important unsolved problem in human genetics. Although these interactions are not directly estimated in the present analysis, the model may provide an approach to their detection. For example, if the variable in question were dental caries, and the observed data were adjusted for fluoride concentration in the drinking water, a reduction in V_{EH} or V_{ES} would identify that factor as an important source of environmental variation. However, if one or more of the genetic variance components is also changed, the result may adumbrate a genetic-environmental interaction.

Following the analysis originally described by FISHER (1918), CROW and KIMURA (1970) and CAVALLI-SFORZA and BODMER (1971) have given formulae which describe the effects of assortative mating on additive genetic variance. However, in practice it may often be difficult, if not impossible, to determine whether assortative mating or common environment is the true cause for an observed correlation between spouses. The present model permits this important distinction to be made. If assortative mating is defined as the tendency for like phenotypes to marry, and is measured by the correlation coefficient,

$$r = \frac{\text{cov}HW_{AS}}{V_T}$$

where $\text{cov}HW_{AS}$ is the husband-wife covariance attributable to assortative mating and V_T the total variance, the expected correlation between the spouses of identical twins because of assortative mating would be r^2. These relationships permit estimation of the expected contribution of assortative mating to the husband-wife covariance and to the covariance between spouses, as shown in Table 3. To measure the effect of assortative mating, one would first estimate the environmental effects over all equations simultaneously. Then, if the observed values of the parental covariances exceed both expected values, a pooled estimate of the variance attributable to assortative mating may be obtained from the formula:

$$V_{AS} = \frac{N_A \Delta_A + N_B \sqrt{V_T \Delta_B}}{N_A + N_B} ,$$

where N_A and N_B are the number of husband-wife and spouse pairs respectively; Δ_A and Δ_B the difference between the observed and expected values for the covariance between husband and wife and between spouses respectively; and V_T is the total variance. If a major effect of assortative mating is detected, the estimates of additive genetic variance could be adjusted appropriately.

As described previously, maternal effects can be detected by contrasting the analysis of variance for the offspring of male and female twins. The effects detected by this analysis would include both prenatal maternal influences and genetically conditioned postnatal differences between mothers in all aspects of mothering and child rearing. Application of the model to the analysis of behavioral trails could have a constructive influence on the current controversy over the sources of variation in mental test scores. The demonstration that maternal

effects make a significant contribution to the total variance might serve to redirect efforts to improve the educational performance of children.

The X chromosome constitutes approximately 5.2% of the haploid female genome (HAMERTON 1971), and consequently, one might expect X-linked genes to make a modest contribution to most quantitative traits. If a substantially greater effect is observed, it could indicate the presence of a polymorphic gene with major effect or a clustering of genes influencing the trait on the X chromosome. It is interesting to speculate as to what effect X inactivation would have on the genetic expression of X-linked genes which influence quantitative traits. Presumably, the dominance component, V_D, represents the cumulative bi-directional effects of dominance summed over many loci. However, in each cell of a normal female, the X-inactivation mechanism acts to create complete dominance of one X-linked allele or the other, and the overall phenotype reflects the summation of these events over all cells of the body. Thus, one might predict that quantitative traits which are influenced to some extent by X-linked genes may show greater evidence of additive effects in males than in females, and more evidence of dominance effects in females than in males, where hemizygosity precludes the possibility of dominance.

Many of the analytic models that have been developed in human genetics are elegant in concept but impractical. KANG *et al.* (1974) considered the efficiency of the monozygotic twin half-sib model and concluded that the variance of estimates of the additive genetic variance begin to stabilize with sample sizes in the range of 100–200 half-sibships for heritabilities of 0.2 or more. These estimates are conservative since they consider only the information provided by the offspring. In any event, the numbers are not inordinately large and since monozygotic twins occur in about 4 per 1,000 births in all racial groups, a large number of suitable families should be available for study in any large population center. It must be acknowledged, however, that the epistatic effects given in Tables 2–4 are highly confounded and the possibility of clearly distinguishing them in any realistic sample size is remote. Because many of the genetic variance components are highly correlated in their effects, the simultaneous inclusion of all variables in the model often leads to large negative estimates of genetic or environmental variance components. This can be interpreted as an indication that the postulated genetic model does not fit the data and a new model which includes fewer variables may then be tried. When several combinations are found which yield only positive estimates of variance components, a chi-square test for goodness of fit may be used to choose between alternative solutions. Other elaborations of the present model are doubtless possible, including a more detailed resolution of maternal effects, direct estimation of the influence of assortative mating, iteration of all variances and covariances in the **V** matrix or possibly the inclusion of genetic-environmental interactions in the model. Further experience with actual data should indicate whether these modifications are necessary or desirable.

A frequent criticism of twin research is that the close relationship of identical twins (ZAZZO 1974) may render conclusions drawn from these studies inapplicable to the general population. In the present model, the two equations of estima-

tion derived from the analysis of variance of the data from the twin parents could, if necessary, be omitted particularly in the *X*-linkage model, where ten equations are available from the offspring data alone. In this way, it would be possible to exploit the unique relationship of identical twins through observations on their offspring, without incurring any of the liabilities that might be associated with the inclusion of data from the twins themselves.

LITERATURE CITED

BENIRSCHKE, K. and C. K. KIM, 1973 Multiple pregnancy. New Eng. J. Med. **288:** 1276–1284.

BULMER, M. G., 1970 *The Biology of Twinning in Man.* Clarendon Press, Oxford.

CAVALLI-SFORZA, L. L. and W. F. BODMER, 1971 Pp. 543–550. In: *The Genetics of Human Populations,* W. N. FREEMAN, San Francisco.

CHRISTIAN, J. C., K. W. KANG, and J. A. NORTON, 1973 Choice of an estimate of genetic variance from twin data. Am. J. Hum. Genet. **26:** 154–161.

COMSTOCK, R., 1955 Theory of quantitative genetics: Synthesis. Cold Spring Harbor Symp. Quant. Biol. **20:** 93–102.

CROW, J. F. and M. KIMURA, 1970 *An Introduction to Population Genetics Theory.* Harper and Row, New York.

ELSTON, R. C., 1975 On the correlation between correlations. Biometrika **62:** 133–140.

FALCONER, D. S., 1961 Pp. 165–185. In: *Quantitative Inheritance,* Roland Press, New York.

FISHER, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinb. **52:** 399–433.

HAMERTON, J. L., 1971 P. 180. In: *Human Cytogenetics.* Academic Press, New York.

HAYMEN, B. J., 1960 Maximum likelihood estimation of genetic components of variation. Biometrics **16:** 369–381.

HORN, J. M., J. C. LOEHLIN and L. WILLERMAN, 1974 Nature-nurture and intelligence: The twin and adoption studies agree. Acta Genet. Med. **23:** 34–35.

JENSEN, A. R., 1974 Kinship correlations reported by Sir Cyril Burt. Behav. Genet. **4:** 1–28.

JINKS, J. C. and D. W. FULKER, 1970 A comparison of the biometrical genetical MAVA and classical approaches to the analysis of human behavior. Psychol. Bull. **73:** 311–349.

KANG, K. W., J. P. LINDEMANN, J. C. CHRISTIAN, W. E. NANCE and J. A. NORTON, 1974 Sampling variances in twin and sibling studies in man. Hum. Hered. **24:** 363–372.

KEMPTHORNE, O., 1957 Pp. 337–339. In: *An Introduction to Genetic Statistics.* John Wiley, New York.

MATHER, K. and J. L. JINKS, 1971 Pp. 225–235. In: *Biometrical Genetics.* 2nd Edition. Cornell Univ. Press, Ithaca.

MILHAM, S., 1964 Pituitary gonadotropin and dizygotic twinning. Lancet. **2:** 566.

MUNSINGER, H., 1975 Children's resemblance to their biological and adopting parents in two ethnic groups. Behav. Genet. **5:** 239–254.

NANCE, W. E., 1976 Genetic studies of the offspring of identical twins. Acta Genet. Med. Gemell. (in press).

NANCE, W. E., M. NAKATA, T. D. PAUL and P. YU, 1974 The use of twin studies in the analysis of phenotypic traits in man. Pp. 23–29. In: *Congenital Defects: New Directions in Research.* Edited by D. T. JANERICH, R. G. SKALKO and I. H. PORTER. Academic Press, New York.

RAO, D. C., N. E. MORTON and S. YEE, 1974 Analysis of family resemblance. II. A linear model for familial correlation. Am. J. Hum. Genet. **26**: 331–359.

SNEDECOR, G. W., 1956 Pp. 271–275. In: *Statistical Methods.* 5th Edition. Iowa State Univ. Press, Ames.

TIZARD, B. and J. REES, 1974 A comparison of the effects of adoption, restoration to the natural mother and continued institutionalization on the cognitive development of four year-old children. Child Dev. **45**: 92–99.

WOLF, B., 1950 Environmental effects in quantitative inheritance. Pp. 81–90. In: *Symposium on Quantitative Inheritance.* Edited by E. C. R. REEVE and C .H. WADDINGTON. Her Majesty's Stationery Office, St. Albans, Great Britain.

ZAZZO, R., 1974 The twin condition and the couple effects on personality development. Acta Genet. Med. Gemell. **23**: 58–59.

Corresponding editor: J. F. CROW