THE ADVANTAGES OF ANALYSING HUMAN VARIATION USING TWINS AND TWIN HALF-SIBS AND COUSINS

C. S. HALEY and KRYSTYNA LAST

Department of Genetics, University of Birmingham, Birmingham B15 2TT

SUMMARY

The "twin family" design is a new strategy for studying quantitative characters in man which overcomes most limitations of earlier designs and which is readily accessible to existing twin research units as it uses only adult twins (both identical and non-identical) and their spouses and offspring and juvenile twins (both identical and non-identical) and their parents. The design yields all information inherent in ordinary twin studies but also permits the simultaneous estimation of more components than any other design. Tests for most genetical and environmental components of variation and the assumptions of the design are provided. Particular advantages are the unambiguous separation of sex-linkage and maternal inheritance, the analysis of the mechanism of assortative mating and the specification of more realistic environmental models. Although several components are confounded the biases are not seriously misleading. However, it would be necessary to include adoption data to resolve the effects of cultural transmission which are otherwise confounded with the family environment. Nevertheless, this design provides a wealth of data on a diversity of relationships and promises to be a valuable tool for the analysis of individual differences in man.

1. INTRODUCTION

THIS paper describes the "twin family"design, an extension of the mono-zygotic twin (MZ) half-sib design of Nance and Corey (1976) which over-comes most shortcomings of the original design. Recent work by Haley, Jinks and Last (1981) has demonstrated how the MZ half-sib design may be used to distinguish sex-linkage and maternal effects in man. The MZ half-sib design involves a study of MZ twins and their spouses and offspring. Independent tests for the detection of sex-linkage and maternal effects are possible using only data from the progeny of the MZ twins when the sexes are considered separately. However, joint estimation of the para-meters used to specify sex-linkage and maternal effects and those specifying autosomal genetic and environmental effects requires the use of data from the parental generation (the MZ twins) and parent-offspring relationships. The use of this additional data for the purpose of estimation is only possible if the stability of genetic and environmental effects across generations is assumed. Furthermore, even after making this assumption, some of the parameters specifying maternal effects remain confounded and with some environmental models, parameters specifying autosomal genetic, environ-mental and maternal effects are confounded (Haley *et al.*, 1981).

In the "twin family" design, the MZ half-sib design of Nance and Corey (1976) is extended to include adult dizygotic (DZ) twins and their spouses and offspring and juvenile MZ and DZ twins and their parents. This design permits the simultaneous estimation of all parameters specifying sex-linkage

and maternal effects and provides a sound basis for the estimation of parameters specifying autosomal genetic and environmental effects utilising data from the juveniles alone. Additionally the full data set provides tests for intergenerational stability of parameters, the representativeness of twins, genotype-environment interactions, genotype-environment covariation, assortative mating, epistasis and sex-limited effects.

2. THE PROPOSED DESIGN

The proposed design comprises adult identical (MZ) twins (male and female pairs), their spouses and offspring, adult non-identical (DZ) twins (male, female and opposite sexed pairs), their spouses and offspring and juvenile twins (male and female MZ and DZ pairs and DZ opposite sexed pairs) and their parents.

An hierarchical analysis of variance of data from the progeny of the adult twins which allows for varying numbers of progeny per family will yield three mean squares:

MS_W: Between progeny within full-sib families.

MS_B: Between the two families of full sibs produced by a pair of twins (these families are genetically related as half-sibs if the parents are MZ twins or as first cousins if the parents are DZ twins).

MS_A: Among families of half-sibs (produced by MZ twins) or cousins (produced by DZ twins).

The three components of variance corresponding to these mean squares are genetically equivalent to, in the case of MZ twin progeny:

σ_w^2: within family variance
σ_b^2: full-sib covariance minus the half-sib covariance
σ_a^2: half-sib covariance

and in the case of DZ twin progeny:

σ_w^2: within family variance
σ_b^2: full-sib covariance minus the first cousin covariance
σ_a^2: first cousin covariance.

The analysis of the data from the twins, both adult and juvenile, is performed by the usual analysis of variance (*e.g.*, Jinks and Fulker, 1971). Further information is provided by the covariances between husbands and wives, an adult twin and their co-twins spouse, the two spouses of a pair of twins, parents and offspring and uncles or aunts and nieces or nephews. The design thus provides a wealth of data on numerous relationships although many of the statistics will be correlated for sampling reasons and because any one individual may contribute to several relationships. In the following discussions it will be assumed that the data are to be analysed using a least squares model fitting approach, making due allowance for the correlations between statistics (Nelder, 1960; Mather and Jinks, 1971; Nance and Corey, 1976). However, the expectations to be given will be equally applicable if some other method of analysis such as maximum likelihood unbalanced pedigree analysis (Lange, Westlake and Spence, 1976; Eaves, Last, Young and Martin, 1978) is utilised.

3. The basic model

The basic model is written in terms of the additive and dominance components of autosomal genetic variation, D_R and H_R respectively (Mather and Jinks, 1971), and additive environmental components of variance. The expectations of all variances and covariances required to construct the expectations of the statistics derived from the analysis are shown in table 1. The basic model makes the simplifying assumption that the overall value of the environmental and genetic components remain constant across generations.

The environmental model is novel in that it allows a fraction of the environmental covariation between individuals to be fixed, for example in childhood, and a fraction to be malleable and responsive to the short term environment. Thus the fixed fraction of environmental covariation may be due to the long term influence of the family environment on the development of a character in the child. At the other extreme the malleable fraction of environmental covariation may simply be an artefact of testing an entire family on the same day. The components representing the malleable part of the environment are marked with an asterisk in table 1. The environmental contribution to variation within families may be made up of fixed and malleable fractions but they are inevitably confounded. The model allows for differing environmental contributions to the covariances between individuals within full-sib families, within half-sib families and within families of cousins. Considering only adult MZ twins and their spouses and offspring the model is equivalent to that of Nance and Corey (1976) if only E_S^*, E_H^* and E_C^* are included and to that of Haley *et al.* (1981) if only E_S, E_H and E_C are included.

All nine parameters in the basic model (*i.e.*, D_R, H_R, E_W, E_S, E_H, E_C, E_S^*, E_H^* and E_C^*) can be estimated simultaneously if all statistics are used. However the model contains many assumptions such as the stability of parameters across generations, the representativeness of twins and the absence of genotype and environmental interactions and covariation. The large body of data provides tests for many of the implicit assumptions in the basic model, the only major exception being the absence of a test for cultural transmission which cannot be detected in any design which does not include adopted individuals. Furthermore, with the failure of the basic model in many cases it is possible to extend the model and provide estimates of the additional parameters.

As always with model fitting it is prudent to follow the principle of Occam's razor and consider initially the simplest models. If simple models fail to provide an adequate explanation of the data then extensions may be explored both by simple scaling tests and model fitting. This process continues until the simplest feasible model which adequately explains the data is arrived at. We strongly advise against the approach of attempting to estimate all possible parameters at the same time unless simpler models have proved inadequate.

4. Examining the assumptions of the basic model

There are two reasons for examining and possibly extending and amending the basic model. Firstly it is always wise to explore alternatives and

TABLE 1

The expectations of the statistics derived from the twin family design in terms of the basic model and the twin environment

					Basic Model						Twin Environment		
		D_R	H_R	E_W	E_S	E_H	E_C	E_S^*	E_H^*	E_C^*	E_{MZ}	E_{DZ}	E_{CH}
General—husband-wife covariance													
Parental Statistics — MZ twins	σ_b^2 between twins	$\frac{1}{2}$	$\frac{1}{4}$	·	1	1	1	1	1	1	1	·	·
	σ_w^2 within twins	1	·	1	·	·	·	·	·	·	1	·	·
	twin-cotwins spouse cov.	·	·	·	1	·	1	1	·	1	·	·	·
	twins spouse-twins spouse cov.	·	·	1	·	1	1	·	1	1	·	·	·
Parental Statistics — DZ twins	σ_b^2 between twins	$\frac{1}{4}$	$\frac{1}{16}$	·	1	1	1	1	1	1	·	1	·
	σ_w^2 within twins	$\frac{1}{4}$	$\frac{3}{16}$	1	·	·	·	·	·	·	·	1	·
	twin-cotwins spouse cov.	·	·	·	1	·	1	1	·	1	·	·	·
	twins spouse-twins spouse cov.	·	·	1	·	1	1	·	1	1	·	·	·
General—parent-offspring cov.													
Intergenerational Statistics — MZ twins	cotwin-offspring cov.	$\frac{1}{4}$	·	·	·	·	1	·	·	1	·	·	·
	cotwins-spouse-offspring cov.	$\frac{1}{4}$	·	·	·	·	1	·	·	1	·	·	·
Intergenerational Statistics — DZ twins	cotwin-offspring cov.	$\frac{1}{8}$	·	·	·	·	1	·	·	1	·	·	·
	cotwins-spouse-offspring cov.	·	·	1	·	·	·	·	·	·	·	·	·
Progeny Statistics — MZ twin progeny	σ_a^2	$\frac{1}{8}$	$\frac{1}{16}$	·	·	1	1	·	1	1	1	·	1
	σ_b^2	$\frac{1}{8}$	$\frac{3}{16}$	·	1	·	·	1	·	·	·	·	·
	σ_w^2	$\frac{1}{4}$	$\frac{3}{16}$	1	·	·	·	·	·	·	·	·	·
Progeny Statistics — DZ twin progeny	σ_a^2	$\frac{1}{16}$	$\frac{1}{16}$	·	·	1	1	·	1	1	·	1	1
	σ_b^2	$\frac{1}{4}$	$\frac{3}{16}$	·	1	·	·	1	·	·	·	·	·
	σ_w^2	$\frac{1}{2}$	$\frac{1}{4}$	1	·	·	·	·	·	·	·	·	·
Progeny Statistics — MZ twins	σ_b^2 between twins	$\frac{1}{4}$	$\frac{1}{16}$	·	1	1	1	·	·	·	1	1	1
	σ_w^2 within twins	$\frac{1}{4}$	$\frac{3}{16}$	1	·	·	·	·	·	·	·	·	·
Progeny Statistics — DZ twins	σ_b^2 between twins	$\frac{1}{4}$	$\frac{1}{16}$	·	1	1	1	·	·	·	1	1	1
	σ_w^2 within twins	$\frac{1}{4}$	$\frac{3}{16}$	1	·	·	·	·	·	·	·	·	·

extensions to a basic model in order to examine the biases introduced into an analysis by the use of inappropriate models. Secondly, they will be required should preliminary investigations reveal the inadequacy of the basic model in practice.

(i) *Intergenerational stability*

A likely cause of the failure of the basic model will be differences in the parameters across generations. This may arise not only because of real biological or environmental changes with age or generation but also because of the use of different scales of measurement in the two generations, a common occurrence with psychological variables. There are no expectations for the form or direction of age or generation effects and their detection will rest upon the failure of the basic model, unless there are gross differences in the phenotypic distributions between generations as revealed by differences in, for example, variances or skewness.

Young, Eaves and Eysenck (1980) have demonstrated one approach to the investigation of intergenerational instability and have shown marked differences in gene action between adults and juveniles for three of the four scales of the Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975). The twin family design contains all the statistics used by Young *et al.* (1980) and more, and thus provides a sound basis for the study of intergenerational stability in the way they described by specifying different parameters for the parental generation, the juvenile generation and intergenerational relationships. Although in theory the consistency of any model over generations may be tested using the method of Young *et al.* (1980), in practice intergenerational instability limits the complexity of models which may be examined, as it greatly increases the ratio of parameters to unique statistics. Therefore, the basic model and its extensions should be examined with the knowledge that data from only a single generation may be available for their resolution. In such circumstances the juvenile generation is the obvious choice as it contains information from the greater number of genetic relationships. We will therefore consider the resolution of the basic model, and later its extensions, using only data from the juveniles.

Utilising only data from the juveniles simplifies the basic model because the fixed and malleable components of environmental variation are no longer separable (*e.g.*, E_S is confounded with E_S^*). Nevertheless the remaining six parameters are not independently estimable from the juvenile data alone for there are only four independent equations for the estimation of the five parameters D_R, H_R, $(E_S + E_S^*)$, $(E_H + E_H^*)$ and $(E_C + E_C^*)$. As a further simplifying assumption dominance could be omitted from the model, in this case the estimates of the remaining parameters become:

$$\hat{E}_W = E_W$$

$$\hat{D}_R = D_R + \tfrac{3}{4} H_R$$

$$(E_S \hat{+} E_S^*) = (E_S + E_S^*) - \tfrac{1}{32} H_R$$

$$(E_H \hat{+} E_H^*) = (E_H + E_H^*) - \tfrac{3}{64} H_R$$

$$(E_C \hat{+} E_C^*) = (E_C + E_C^*) - \tfrac{3}{64} H_R.$$

Thus, in the presence of dominance, the underestimation of the environmental components is very small and will not lead to a serious overestimate of the magnitude of genetic variation. If any of the environmental estimates are zero or negative they may be omitted and an alternative model including dominance explored. When all of $(E_S + E_S^*)$, $(E_H + E_H^*)$ and $(E_C + E_C^*)$ are significantly positive, it may be possible to detect dominance by tests independent of model fitting procedures (Martin et al., 1978).

An alternative simplification of the basic model is to assume that the environmental contribution to the covariance between half-sibs is the same as that to the covariance between cousins (i.e., $(E_H + E_H^*) = 0$), the parameter estimates then become

$$\hat{E}_W = E_W$$

$$\hat{D}_R = D_R + 16(E_H + E_H^*)$$

$$\hat{H}_R = H_R - \tfrac{64}{3}(E_H + E_H^*)$$

$$(E_S \,\hat{+}\, E_S^*) = (E_S + E_S^*) - \tfrac{2}{3}(E_H + E_H^*)$$

$$(E_C \,\hat{+}\, E_C^*) = (E_C + E_C^*) - (E_H + E_H^*).$$

Thus unless $(E_H + E_H^*)$ is very small it will result in negative estimates of H_R and the model will be rejected. Therefore neither assuming $H_R = 0$ nor assuming $(E_H + E_H^*) = 0$ will seriously mislead us if our assumption is erroneous.

(ii) The twin environment

Critics of the twin method have often argued that the covariance between MZ twins is enhanced compared to that between DZ twins by the greater similarity of their environment. This model can simply be accommodated by the inclusion of a parameter E_{MZ} in the expectations for covariances between MZ twins and in the expectations for the within family variance of all other types of family as shown in table 1. With data from only twins reared together the model E_{MZ}, E_W and $(E_S + E_S^* + E_H + E_H^* + E_C + E_C^*)$ is simply a re-parameterisation of the gene-environment model. However, with data from additional relatives it becomes possible to estimate simultaneously all parameters of the basic model plus E_{MZ} if our assumption of intergenerational stability holds. With intergenerational instability no simple test for the presence of E_{MZ} is available and its undetected presence may lead to the inflation of estimates of genetic parameters, for example using juvenile data only:

$$\hat{E}_W = E_W$$

$$\hat{D}_R = D_R + \tfrac{3}{4}H_R + 4E_{MZ}$$

$$(E_S \,\hat{+}\, E_S^*) = (E_S + E_S^*) - \tfrac{1}{32}H_R - \tfrac{1}{2}E_{MZ}$$

$$(E_H \,\hat{+}\, E_H^*) = (E_H + E_H^*) - \tfrac{3}{64}H_R - \tfrac{1}{4}E_{MZ}$$

$$(E_C \,\hat{+}\, E_C^*) = (E_C + E_C^*) - \tfrac{3}{64}H_R - \tfrac{1}{4}E_{MZ}.$$

Thus the estimate of the genetical variation, $D_R + \tfrac{3}{4}H_R$, is inflated by E_{MZ} whilst the family environment components are deflated by H_R and E_{MZ}

and may become non-significant or even negative. However, we can be more optimistic than this, for if we consider a purely environmental model for the juvenile data we have:

$$\hat{E}_W = E_W$$

$$(E_S \hat{+} E_S^*) = (E_S + E_S^*) + \tfrac{1}{8}D_R + \tfrac{1}{16}H_R$$

$$(E_H \hat{+} E_H^*) = (E_H + E_H^*) + \tfrac{1}{16}D_R$$

$$(E_C \hat{+} E_C^*) = (E_C + E_C^*) + \tfrac{1}{16}D_R$$

$$\hat{E}_{MZ} = E_{MZ} + \tfrac{1}{4}D_R + \tfrac{3}{16}H_R.$$

As all parameter estimates where confounded are inflated by the genetical effects, any non-significant parameters may be omitted thus allowing estimation of D_R or D_R and H_R. In the absence of $(E_H + E_H^*)$ we have

$$\hat{E}_W = E_W$$

$$\hat{D}_R = D_R$$

$$(E_S \hat{+} E_S^*) = (E_S + E_S^*) + \tfrac{1}{16}H_R$$

$$(E_C \hat{+} E_C^*) = (E_C + E_C^*)$$

$$\hat{E}_{MZ} = E_{MZ} + \tfrac{3}{16}H_R.$$

The model outlined above can be extended to allow for the possibility that DZ twins share a more similar environment than ordinary full-sibs. This is accomplished by the inclusion of the term E_{DZ} in the expectations shown in table 1, E_{DZ} can be estimated independently of all other parameters from the progeny data alone.

All DZ twins have separate chorions (*i.e.*, are dichorionic) but whereas some MZ twins have separate chorions, the majority of pairs share a single chorion (*i.e.*, are monochorionic). Monochorionic MZ twins are less variable than dichorionic MZ twins for some traits, for example cord blood cholesterol level (Corey *et al.*, 1976) and fingertip dermatoglyphics (Reed *et al.*, 1978). In this case the model can be extended by the inclusion of the parameter E_{CH} which contributes to the expectations for all within family variances except monochorionic MZ twins. In the absence of the obstetrical data required to diagnose monochorionic and dichorionic twins, differences in the total variance of MZ twins and other inviduals may indicate a chorionic effect. This may be accommodated in the model by the inclusion of E_{CH}, which is independently estimable with data from either generation, in the expectations for all within family variances except that for MZ twins as shown in table 1.

(iii) *Genotype-environment interaction* $(G \times E)$

Eaves *et al.* (1977) have demonstrated that the power of the test for the detection of $G \times E$ from model fitting is low. Further, the possibility of the resolution of $G \times E$ via a model fitting approach will depend upon the complexity of the other parameters in the model. However, Jinks and Fulker (1971) and Eaves *et al.* (1977) have shown how systematic $G \times E$ may be detected by the regression of the intra-pair differences of MZ twins on their mean. Furthermore Martin (1977) demonstrates that this is a

powerful test. A less powerful approach suggested by Eaves *et al.* (1977) is the investigation of the distributional skewness of the data. However, skewness is produced by all forms of genetical non-additivity and by unequal gene frequencies in the presence of only additive genetic variation. Once $G \times E$ has been detected the usual approach is to remove it by rescaling the data, simple transformations being often all that is necessary. However, if it is desirable to retain the original scale or rescaling the data does not remove the $G \times E$, then it is possible to incorporate $G \times E$ in the model, although its effects are likely to be confounded with other parameters except in the simplest of models. This is especially so in the absence of adoption data. Nevertheless, Eaves *et al.* (1977) suggest that the biases introduced by undetected $G \times E$ are unlikely to be major.

(iv) *Genotype-environment covariation (Cov GE)*

Three mechanisms leading to covariance between genotype and environment have been recognised. These are individuals choosing their own environment, sibling interactions and cultural transmission.

When individuals choose their own environment and the choice is partly determined by their genotype such that the genotypic variation is inflated or deflated, then covariation is established between genotype and environment. This effect is inevitably completely confounded with direct genotypic effects. Indeed, as it can be regarded as an extension of the genotype, it may be considered that it does not introduce any serious bias into any inferences drawn from the analysis (Jinks and Fulker, 1971).

Jinks and Fulker (1971) have shown how the genotype-environment covariance produced by sibling competition may be included in the analysis of twin data. Eaves (1976a) has produced a general model for interaction between siblings (sibling effects) and shows that sibling competition leads to a reduction of the total variance of MZ twins compared with DZ twins and ordinary siblings whereas sibling co-operation will have the opposite effect. Note that sibling competition could be mistaken for a chorionic effect as it reduces the variance of MZ twins relative to other types of siblings. However, the twin family design provides a means of distinguishing between sibling effects and a chorionic effect through a comparison of the variances of families of different size. Sibling effects cannot contribute to the variance of singletons and so in presence of sibling competition we can predict the hierarchy of total variances: singletons > DZ twins > MZ twins and in the case of sibling co-operation we expect the hierarchy MZ twins > DZ twins > singletons. If these comparisons detect sibling effects, it would be possible to include them in the model. However, the model of Eaves (1976a) would need to be extended to families of size greater than two.

The third mechanism that has been recognised by which genotype-environment covariation may be generated is cultural transmission as defined by Cavalli-Sforza and Feldman (1973) and Eaves (1976b). The best test for cultural transmission is a comparison of the total variances of adopted individuals and individuals reared by their natural parents (Eaves *et al.*, 1977). Although there is no specific test for cultural transmission in the twin family design, its effects as defined by Eaves (1976b) are confounded with the environmental components $(E_S + E_S^*)$, $(E_H + E_H^*)$ and

$(E_C + E_C^*)$, and do not inflate D_R, as estimated from the juvenile data. With the complete data set, cultural transmission may result in the failure of the basic model. If it does not, it is likely to remain largely confounded with the parameters specifying the effects of the common environment and not seriously bias our estimates of direct genetic effects.

(v) *Assortative mating*

Twins and their spouses provide a good basis for the detection of assortative mating and the analysis of its foundation (Nance and Corey, 1976; Eaves, 1979). It is possible to distinguish marital correlations caused by assortative mating from those due to the effects of the common environment shared by spouses by a comparison of the correlation between a twin (either MZ or DZ) and their co-twins' spouse and the correlation between the two spouses of the twins. Under the model of Eaves (1979), these two correlations are expected to be the same if due to the effects of the common environment but differ if due to assortative mating. With the detection of assortative mating a comparison of the correlation between the spouses of MZ twins with that between the spouses of DZ twins may reveal any genetic basis for assortation. Additionally, Eaves (1979) demonstrates that, if assortation is based solely on the basis of the measured phenotype, then the correlation between a twin and their co-twins spouse is expected to be the product of the twin correlation and the marital correlation, and the correlation between the two spouses of twins is expected to be the product of the twin correlation and the marital correlation squared. When assortation is based upon the measured phenotype, the basic model may be amended by the inclusion of the appropriate expectations (Fisher, 1918; Eaves, 1975).

When assortative mating is asymmetric (*i.e.*, females choose different qualities in males than males choose in females), Eaves (1979) demonstrates that the expected pattern of correlations between the spouses of twins differs depending upon the sex of the twins. Additionally this model predicts that in like sexed twins there will be no correlations between twins and their co-twins' spouses or between husbands and wives. However, continued assortation will generate gametic disequilibrium between genes controlling the two characters and may produce significant values for these correlations (Eaves and Heath, 1981). Asymmetric assortative mating may produce maternal half-sib covariances greater than paternal half-sib covariances thus mimicking maternal effects (Nance, 1979; Eaves and Heath, 1981), although it does not produce maternal parent-offspring covariances larger than paternal parent-offspring covariances. This last fact and a comparison of the correlations between the spouses of twins allows the detection of asymmetric assortative mating and its separation from maternal effects.

(vi) *Epistasis*

The expected contribution of epistasis between pairs of loci to the statistics derived from the twin family design are shown in table 2 in terms of I_R, J_R and L_R (Mather, 1974). I_R represents the effect of interactions between pairs of homozygous loci, J_R represents the effect of interactions

TABLE 2

The expectations of the genetic relationships in the twin family design in terms of the components of two locus epistasis

	I_R	J_R	L_R
Total variance	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$
Full-sib covariance	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{256}$
Half-sib covariance	$\frac{1}{64}$	.	.
Cousin covariance	$\frac{1}{256}$	.	.
Parent-offspring covariance	$\frac{1}{16}$	.	.
Avuncular covariance	$\frac{1}{64}$	.	.

between pairs of loci, one heterozygous and one homozygous and L_R represents the effect of interactions between pairs of heterozygous loci. These three components cannot be estimated with the data from a single generation. With the progeny data alone undetected epistasis mainly biases estimates of H_R and thus any parameters with which H_R is confounded. The complete data set enables the estimation of D_R, E_W, E_S, E_H, E_C, E_S^*, E_H^*, E_C^* and I_R simultaneously, however, only two of H_R, J_R and L_R can be estimated. Estimates of I_R, J_R and L_R will have large sampling variances in most typical studies and are unlikely to be significant. If they remain undetected, they inflate the estimate of H_R which can, therefore, be considered as an estimate of general non-additive genetic variation.

(vii) *Sex-limited autosomal variation*

If the data are divided into male and female sets, separate models with male and female specific components can be applied to the two sexes. Furthermore, we can follow the example of Eaves (1977) and, after the removal of the mean sex difference, specify further components for the covariance of opposite-sexed individuals. For example, the additive component of genetical variation D_R, which represents genes common to both sexes, can be replaced by D_{Rm} in males, by D_{Rf} in females and by D_{Rmf} in expectations involving covariances between pairs of opposite-sexed individuals. This model can be extended to include further genetic components and different within and between family environmental components within and between sexes. The contribution of sex-limited autosomal additive and dominance variation to the basic statistics of the design is given in table 3. Separation of the data by sex, increases the number of statistics and, therefore, the large number of parameters required to specify sex-limited variation are no more confounded than are those in the basic model.

(viii) *Sex-linked variation*

The model for sex-linked variation of Mather and Jinks (1963) can be extended to include all statistics in the twin family design. The model uses three parameters to define the additive effects of sex-linked loci in males, in females and in covariances between males and females. These are D_B, D_S and D_U respectively. An additional parameter, H_X, specifies the dominance effects of sex-linked loci in females. These four parameters are

as defined in Mather and Jinks (1963) and in Haley *et al.* (1981). However, note that $D_B = D_S = D_U$ only in the absence of dominance and with no sex-limitation (*i.e.*, complete dosage compensation) of the sex-linked loci. The contribution of these four parameters to the expectations of the basic statistics in the twin family design is shown in table 3. Note that in some circumstances sex-linked variation and sex-limited autosomal variation may be confused by a comparison of full-sib correlations, this occurs when $D_{Rm} > D_{Rmf} < D_{Rf}$. Comparisons of parent-offspring and half-sib correlations would be unlikely to lead to this confusion.

(ix) *Maternal effects*

The models for genotypic maternal effects given in Haley *et al.* (1981) can be expanded to encompass all of the basic statistics derived from the twin family design. These models are appropriate where the maternal effects have some genetic basis. Haley *et al.* (1981) propose two models for genotypic maternal effects, the one character model, in which the same loci in mother and offspring influence the phenotype under study in the offspring, and the two character model, in which the maternal loci responsible for the maternal effect are not those which directly influence the development of the phenotype in the offspring. Four parameters are required to specify the genotypic maternal effects of the one character model. These are MD_R and MH_R for the additive and dominance maternal effects respectively of the maternal gene, DMD_R for the covariance between the direct and maternal additive effects and HMH_R for the covariance between direct and maternal dominance effects. Only MD_R and MH_R are required to specify the expectations of the two character model in the absence of gametic disequilibrium between the two sets of loci. The expectations are given in table 3 in terms of the four parameters required to specify the one character model. Omission of DMD_R and HMH_R gives the expectations derived from the two character model.

Asymmetry is introduced by maternal effects into the various statistics derived from the twin family design. However, this asymmetry is only between the covariance of maternal relatives and the covariance of paternal relatives. There is no asymmetry between male and female progeny variances. If statistics are not averaged over sexes, sex-linked variation, or sex-limited autosomal variation, or a combination of both will lead to sexual asymmetry and are thus unlikely to be mistaken for maternal effects. However, if doubt exists as to which model (sex-limited autosomal variation, sex-linked variation or maternal effects) or combination of models is appropriate, the progeny data alone are sufficient to enable the estimation of D_S, D_U, D_B, H_X, MD_R, MH_R and DMD_R in addition to those parameters which can be estimated in the basic sex-limited autosomal model.

5. DISCUSSION

The classical twin study has shown that for many traits, MZ twins are substantially more alike than DZ twins, suggesting that genetic factors may contribute to individual differences. However, unsupported twin studies are open to many criticisms. The singularity of the twin condition leads to doubts regarding the applicability of conclusions derived from twin

TABLE 3

The expectations of the genetic relationships in terms of the components of sex-associated variations

	D_{Rf}	H_{Rf}	D_{Rmf}	H_{Rmf}	D_{Rm}	H_{Rm}	D_S	D_U	D_B	H_X	MD_R	MH_R	DMD_R	HMH_R
Total variance (female)	$\frac{1}{2}$	$\frac{1}{4}$	.	.	.	.	$\frac{1}{2}$	.	.	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	.
Total variance (male)	.	.	.	.	$\frac{1}{2}$	$\frac{1}{4}$	.	.	1	.	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	.
Sister-sister cov.	$\frac{1}{4}$	$\frac{1}{16}$	.	.	.	.	$\frac{3}{8}$	.	.	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	.
Sister-brother cov.	.	.	$\frac{1}{2}$	$\frac{1}{16}$	.	.	.	$\frac{1}{4}$	.	.	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	.
Brother-brother cov.	.	.	$\frac{1}{4}$	.	$\frac{1}{4}$	.	.	.	$\frac{1}{2}$	.	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	.
Mother-daughter cov.	$\frac{1}{4}$	.	$\frac{1}{4}$	.	.	.	$\frac{1}{4}$	.	.	.	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{5}{8}$	$\frac{1}{8}$
Mother-son cov.	.	.	$\frac{1}{4}$	.	.	.	.	$\frac{1}{2}$	.	.	$\frac{1}{4}$	.	$\frac{5}{8}$	$\frac{1}{8}$
Father-daughter cov.	.	.	$\frac{1}{4}$	.	.	.	.	$\frac{1}{2}$	.	.	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$
Father-son cov.	.	.	.	.	$\frac{1}{4}$	.	.	.	$\frac{1}{2}$	.	.	.	$\frac{1}{2}$	$\frac{1}{8}$
Female MZ twin families	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Female half-sib cov.	$\frac{1}{8}$	.	.	.	.	.	$\frac{1}{8}$	.	.	.	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	.
Female-male half-sib cov.	.	.	$\frac{1}{8}$	.	.	.	.	$\frac{1}{4}$	.	.	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	.
Male half-sib cov.	.	.	.	.	$\frac{1}{8}$	.	.	.	$\frac{1}{2}$	.	.	.	.	.
Male MZ twin families	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Female half-sib cov.	$\frac{1}{8}$	.	.	.	.	.	$\frac{1}{4}$	.	.	.	.	.	.	.
Female-male half-sib cov.	.	.	$\frac{1}{8}$	.	.	.	.	.	.	.	.	.	.	.
Male half-sib cov.	.	.	.	.	$\frac{1}{8}$	.	.	.	.	.	.	.	.	.

Female DZ twin families

Female cousin cov.	$\frac{1}{16}$
Female-male cousin cov.	$\frac{1}{16}$
Male cousin cov.	$\frac{1}{16}$
Aunt-niece cov.	$\frac{1}{8}$
Aunt-nephew cov.	$\frac{1}{16}$

Male DZ twin families

Female cousin cov.	$\frac{1}{16}$
Female-male cousin cov.	$\frac{1}{16}$
Male cousin cov.	$\frac{1}{8}$
Uncle-niece cov.	$\frac{1}{16}$
Uncle-nephew cov.	$\frac{1}{8}$

Female-male DZ twin families

Female cousin cov.	$\frac{1}{16}$
Daughter of female twin-son of male twin cov.	$\frac{1}{16}$
Daughter of male twin-son of female twin cov.	$\frac{1}{16}$
Male cousin cov.	$\frac{1}{16}$
Aunt-niece cov.	$\frac{1}{8}$
Aunt-nephew cov.	$\frac{1}{8}$
Uncle-niece cov.	$\frac{1}{16}$
Uncle-nephew cov.	$\frac{1}{8}$

studies to the general population and the environmental similarity and close relationship of MZ twins even undermines the conclusion that genetic variability exists in the twin population. Furthermore, the classical twin study yields only three independent statistics and thus many causes of variation are inextricably confounded.

The move towards the study of extended data sets involving many relationships has been gradually accelerating over recent years because of the drawbacks of the classical twin study. One possibility is the extension of the classical twin study to include the parents of twins. However, Young *et al.* (1980) have demonstrated that studies involving several generations may well be faced with the problem of intergenerational instability.

The MZ half-sib design (Nance and Corey, 1976) is a further step towards more general designs. However, for estimation purposes, data are required from MZ twin parents, their progeny and intergenerational relationships. Thus problems of intergenerational instability and the representativeness of twins are confounded and undermine the value of the estimates obtained.

The twin family design represents a logical extension of previous twin designs as it involves relationships all of which could be available to any centre of twin research as they are based solely upon twin families. Not only does the twin family design involve readily available relationships, it includes all of the most powerful natural contemporary relationships in a single generation. Furthermore, it provides the opportunity to investigate intergenerational stability and the representativeness of twins and hence conclusions drawn from the design can be held with more confidence, than those based on unsupported twin studies.

The complete data set from the twin family design allows the independent estimation of many parameters hitherto confounded. Even in the presence of intergenerational instability the separate generations provide very useful sets of equations. The juvenile data are particularly useful for investigating sex-associated variation and the representativeness of twins. The intergenerational relationships allow the estimation of common environmental effects with which cultural transmission may be confounded. Of especial interest in the parental data are the effects of assortative mating.

We would be foolish to suggest that any set of data, which it is practical to collect, would provide enough information to allow the estimation of more than a limited number of parameters. It is theoretically possible to estimate simultaneously more than 20 parameters with the complete data set from the twin family design. Unfortunately, if the evaluation of this number of parameters were actually attempted it is possible that none would have estimates significantly greater than zero. This of course is a problem common to any design, for with model fitting procedures, as the number of parameters estimated increases so does the variance of those estimates. However, this does not undermine the usefulness of the twin family design, for in any set of data, only a few effects will contribute an appreciable proportion of the variance. One major strength of the design is that it enables the independent estimation of parameters of alternative hypotheses and thus the evaluation of those hypotheses. However, in some instances the use of individual scaling tests may be the more powerful method of detecting an effect. Model fitting then allows the estimation of parameters and the evaluation of rival hypotheses.

We do not wish to argue that the twin family design provides the best of all possible worlds, or that any extension of it would be worthless. Far from it, the design might be usefully extended to include adoptees and their adopting and natural families in the investigation of cultural transmission or phenotypic maternal effects, and other extensions are possible. However, the design does include all the most readily available and powerful natural relationships from the human population and has great potential for the elucidation of the causes of individual differences in both behavioural and physical traits. We believe that the twin family design is one of the most comprehensive and economical yet suggested for analysing individual differences in man.

6. REFERENCES

CAVALLI-SFORZA, L. L., AND FELDMAN, M. W. 1973. Cultural versus biological inheritance: phenotypic transmission from parents to children (a theory of the effect of parental phenotypes on children's phenotypes). *Am. J. Human Genet., 25,* 618-637.

COREY, L. A., KANG, K. W., CHRISTIAN, J. C., NORTON, J. A., HARRIS, R. E., AND NANCE, W. W. 1976. Effects of chorion type on variation in cord blood cholesterol of monozygotic twins. *Am. J. Human Genet., 28,* 433-441.

EAVES, L. J. 1975. Testing models for variation in intelligence. *Heredity, 34,* 132-136.

EAVES, L. J. 1976a. A model for sibling effects in man. *Heredity, 36,* 205-214.

EAVES, L. J. 1976b. The effect of cultural transmission on continuous variation. *Heredity, 37,* 41-57.

EAVES, L. J. 1977. Inferring the causes of human variation. *J. Roy. Statist. Soc. Ser. A., 140,* 324-355.

EAVES, L. J. 1979. The use of twins in the analysis of assortative mating. *Heredity, 43,* 399-409.

EAVES, L. J., AND HEATH, A. C. 1981. On the detection of asymmetric assortative mating. *Nature, 289,* 205-206.

EAVES, L. J., LAST, K. A., MARTIN, N. G., AND JINKS, J. L. 1977. A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *Brit. J. Math. Statist. Psychol., 30,* 185-197.

EAVES, L. J., LAST, K. A., YOUNG, P. A., AND MARTIN, N. G. 1978. Model-fitting approaches to the analysis of human behaviour. *Heredity, 41,* 249-320.

EYSENCK, S. G. B., AND EYSENCK, H. J. 1975. *Manual of the EPQ (Eysenck Personality Questionnaire).* University of London Press, London.

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

HALEY, C. S., JINKS, J. L., AND LAST, K. A. 1981. The monozygotic twin half-sib method for analysing maternal effects and sex-linkage in humans. *Heredity, 46,* 227-238.

JINKS, J. L., AND FULKER, D. W. 1971. Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychol. Bull., 73,* 311-349.

LANGE, K., WESTLAKE, J., AND SPENCE, M. A. 1976. Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann. Human Genet.* (London), *39,* 485-491.

MARTIN, N. G. 1977. The classical twin study in human behaviour genetics. Ph.D. Thesis, University of Birmingham.

MARTIN, N. G., EAVES, L. J., KEARSEY, M. J., AND DAVIES, P. 1978. The power of the classical twin study. *Heredity, 40,* 97-116.

MATHER, K. 1974. Non-allelic interactions in continuous variation of randomly breeding populations. *Heredity, 32,* 414-419.

MATHER, K., AND JINKS, J. L. 1963. Correlations between relatives arising from sex-linked genes. *Nature, 198,* 314-315.

MATHER, K., AND JINKS, J. L. 1971. *Biometrical Genetics*, 2nd Ed. Chapman and Hall, London.

NANCE, W. E. 1979. In *Genetic Analysis of Common Diseases: Applications to Predicitive Factors in Coronary Disease*, eds. C. F. Sing and M. Skolnick. Alan Liss, New York.

NANCE, W. E., AND COREY, L. A. 1976. Genetic models for the analysis of data from the families of identical twins. *Genetics, 83*, 811-825.

NELDER, J. A. 1960. The estimation of variance components in certain types of experiment on quantitative genetics. In *Biometrical Genetics* Ed. O. Kempthorne. Pergamon Press, London.

REED, T., UCHIDA, I. A., NORTON, J. A., AND CHRISTIAN, J. C. 1978. Comparisons of dermatoglyphic patterns in monochorionic and dichorionic monozygotic twins. *Am. J. Human. Genet., 30*, 383-391.

YOUNG, P. A., EAVES, L. J., AND EYSENCK, H. 1980. Intergenerational stability and change in the causes of variation in personality. *Person. Ind. Diff. 1*, 35-55.