

# A Model-Fitting Implementation of the DeFries-Fulker Model for Selected Twin Data

S. Purcell<sup>1,2</sup> and P. C. Sham<sup>1</sup>

Received 20 Dec. 2001—Final 23 Aug. 2002

In this research note, DeFries-Fulker (DF) regression analysis is reframed in model-fitting terms, where an individual's expected score is modeled as a function of their co-twin's proband status. This more flexible implementation of the DF model allows DZ-O twins to be incorporated in a sex-limitation model. Brief simulation results are presented along with the Mx scripts used.

**KEY WORDS:** Twins; DeFries-Fulker analysis; sex differences; maximum likelihood estimation.

DeFries and Fulker (1985, 1988) proposed a method for analyzing twin data based on a simple multiple regression model in which a cotwin's score is predicted from a proband score and the twin-pair's coefficient of genetic relatedness. This model is particularly suited to twin data ascertained for extreme scores. Alternatively, probands may be defined as extreme scorers in a larger unselected sample. In this way, the etiology of deviant scores can be addressed, as opposed to that of individual differences over the normal range. In this research note, the basic DF model is reframed in model-fitting terms. One advantage of this reframing is illustrated in the easy inclusion of opposite-sex DZ twins in a sex-limited analysis. Other extensions should also be possible, including bivariate analysis.

For twin pair  $i$ , a proband with score  $P_i$  is selected if  $P_i$  is greater than (or less than, depending on the direction of selection) a threshold  $t$ . The original DF model predicts the score of the co-twin  $C_i$  as a function of  $P_i$  and  $R_i$ , the coefficient of genetic relatedness for pair  $i$ , coded 1 for MZ twins and 0.5 for DZ twins:  $C_i = b_0 + b_1P_i + b_2R_i + \varepsilon_i$ . The coefficient  $b_2$  is therefore an estimate of heritability based on the differential regression to the mean of MZ versus DZ co-twins, providing the data are transformed such that the proband

mean is 1 and the population mean is 0. In this case,  $b_2$  is equal to twice the difference between MZ and DZ transformed co-twin means, as these reflect the twin correlations. Pairs with two probands are double-entered; pairs with no probands are excluded from analysis.

This model can be easily specified in model-fitting terms: an individual's expected score is a function of their co-twin's proband status. Each observation in the raw datafile corresponds to a twin pair, rather than a proband-cotwin pairing, and contains the following information: zygosity (coded 1/2 for MZ/DZ), continuous trait score for twin 1 and twin 2, and proband status for twin 1 and twin 2 (coded 0/1). Prior to analysis, the data must be transformed in such a way that an individual's score  $x$  becomes  $(x - \bar{x}_0)/(\bar{x}_1 - \bar{x}_0)$  where  $\bar{x}_0$  is the population mean and  $\bar{x}_1$  is the proband mean. These means are typically zygosity-specific, although zygosity non-specific means should be acceptable too.

For transformed scores, the expected means vector  $[m_1 \ m_2]$  is  $[a^2 + c^2] \otimes \mathbf{P}$  for MZ twins and  $[a^2/2 + c^2] \otimes \mathbf{P}$  for DZ twins, where  $\mathbf{P} = [p_2 \ p_1]$  and  $p_j$  is 0 if twin  $j$  is not a proband or 1 if twin  $j$  is a proband. The parameters  $a^2$  and  $c^2$  represent proportions of variance in the general population attributable to additive genetic and shared environmental effects respectively; nonshared environmental variance is  $e^2 = 1 - a^2 - c^2$ . The covariance structure for both MZ and DZ twins is

$$\begin{bmatrix} v + dp_2 & 0 \\ 0 & v + dp_1 \end{bmatrix}.$$

<sup>1</sup> Social, Genetic & Developmental Research Centre, Institute of Psychiatry, London, SE5 8AF, UK.

<sup>2</sup> To whom correspondence should be addressed. e-mail: s.purcell@iop.kcl.ac.uk

The parameter  $v$  is the residual variance; the parameter  $d$  accounts for the difference in variance between individuals with proband cotwins and those without. An Mx script to implement this model is presented in Appendix A. Boundaries and confidence intervals on  $a_2$  and  $c_2$  can be implemented within Mx, as well as other benefits of a model-fitting approach.

In the regression method, the standard error of the  $b_2$  estimate must be modified to account for the double-entry procedure by multiplying by a factor of  $\sqrt{N/n}$  where  $N$  is the number of probands and  $n$  is the number of pairs with at least one proband. A  $\chi^2_1$  statistic may be obtained using the Wald test,  $(b_2/SE_{b_2})^2$ . In the model-fitting method, the likelihood ratio statistic for a model with  $a^2$  estimated versus a submodel with  $a^2$  fixed to 0 provides a test of genetic influence. A  $\chi^2_1$  statistic is calculated as minus twice the difference in log-likelihoods.

Table I clarifies the differences in data structures under the two methods. For the original DF method ( $P$  is proband score,  $C$  is co-twin score) note that twin pair 2 is entered twice (both members are probands) whereas twin pair 3 is not entered at all (neither member is a proband). In contrast, for the model-fitting method, the data is entered as in the original file, only with proband status variables appended. Note that before analysis these data would also have to be transformed. Utility scripts used in the preparation of datasets can be obtained from the website <http://statgen.iop.kcl.ac.uk/df/>.

## SIMULATION

Eights models were simulated 50 times each as shown in Table II under two *unselected* sample sizes: 1,000 MZ and 1,000 DZ pairs; 5,000 MZ and 5,000 DZ pairs. The genetic effect increases from 0% to 75% in 25% intervals for the first four models; the rest of the variance is nonshared between twins. The next four mod-

**Table I.** Data Formats for DF Regression and Model-Fitting Methods: Probands Defined as Scoring Greater Than or Equal to 7

Raw data				Original DF				Model-fitting DF					
ID	Zyg	T1	T2	ID	Zyg	P	C	ID	Zyg	T1	T2	P1	P2
1	MZ	8	5	1	MZ	8	5	1	MZ	8	5	1	0
2	MZ	9	7	2	MZ	9	7	2	MZ	9	7	1	1
3	DZ	4	5	2	MZ	7	9	3	DZ	4	5	0	0
4	DZ	2	8	4	DZ	8	2	4	DZ	2	8	0	1

Scores would also have to be transformed prior to analysis. For the original DF,  $P$  is the proband score and  $C$  is the co-twin score; note how twin pair 2 is double-entered and twin pair 3 is excluded.

**Table II.** Basic DF Simulation Models

Model	$a^2$	$c^2$	$e^2$
1	0.000	0.000	1.000
2	0.250	0.000	0.750
3	0.500	0.000	0.500
4	0.750	0.000	0.250
5	0.000	0.500	0.500
6	0.250	0.375	0.375
7	0.500	0.250	0.250
8	0.750	0.125	0.125

els are similar, except that half of the non-genetic variance is shared between twins. Individuals scoring in approximately the top 5% on the normally-distributed continuous trait were designated as probands, using a threshold of 1.64485 standard deviation units. Therefore, for the  $N = 2,000$  samples, we would expect 50 probands of each zygosity; for the  $N = 10,000$  samples, we would expect 250 probands of each zygosity.

Table III presents the simulation results for the basic model. As expected, the model-fitting and regression approaches give identical parameter estimates for  $a^2$  and  $c^2$  and so these are not shown separately. In practice, it might be desirable to place boundaries on these parameters (e.g., estimating  $a$  rather than  $a^2$ ).

In general, power to detect a genetic effect using the DF model on a selected sample will always be much lower than for individual differences analysis performed on the entire sample. That is, the DF model would only be used when the proband-ascertained structure of the sample dictates, or when the question of deviant versus normal scores is of interest. If only the selected sample were available, individual differences model-fitting could not be straightforwardly applied without correction for ascertainment, and would probably not offer greater power in this case.

The DF likelihood ratio test statistic appears to be equivalent to the corrected Wald test. For models 1 and 5, both with no genetic component, we would expect an average  $\chi^2_1$  value of 1.0 if the tests are valid under the null. Although we observe values higher than expected under the null, it is common to both methods and, more importantly, only seen for the smaller sample size—*asymptotically*, both tests appear not to be liberal. Under non-null simulation models, the average  $\chi^2_1$  always exceeds the critical value associated with a 5% type I error rate and one degree of freedom (3.84146). The test statistics from the simulations with 10,000 pairs are reproduced in Figure 1 to show their overlap. It is possible that the regression method, based

**Table III.** Basic DF Model Simulation Results: Full Sample N = 2,000 and N = 10,000 Pairs

Model	N = 2,000					N = 10,000				
	$a^2$	$c^2$	$e^2$	LRT	Wald	$a^2$	$c^2$	$e^2$	LRT	Wald
1	0.013	0.000	0.987	2.07	2.03	0.003	-0.003	1.000	1.05	1.02
2	0.247	-0.001	0.753	4.52	4.33	0.267	-0.012	0.745	20.02	19.26
3	0.502	-0.008	0.506	15.81	15.27	0.503	0.000	0.497	76.73	74.07
4	0.754	-0.002	0.248	40.95	41.60	0.750	-0.001	0.251	196.12	194.94
5	0.035	0.471	0.494	1.41	1.29	-0.003	0.503	0.500	0.81	0.75
6	0.259	0.365	0.376	5.50	5.14	0.243	0.378	0.379	22.78	21.01
7	0.508	0.237	0.255	20.17	19.35	0.504	0.245	0.251	100.35	94.99
8	0.784	0.088	0.129	52.49	53.42	0.737	0.139	0.124	228.67	229.26

Top 5% selected as probands; likelihood ratio test (LRT) gives average  $\chi^2$  difference obtained by fixing  $a^2$  to 0; Wald test statistic gives test of  $a^2$  from regression-based DF method.

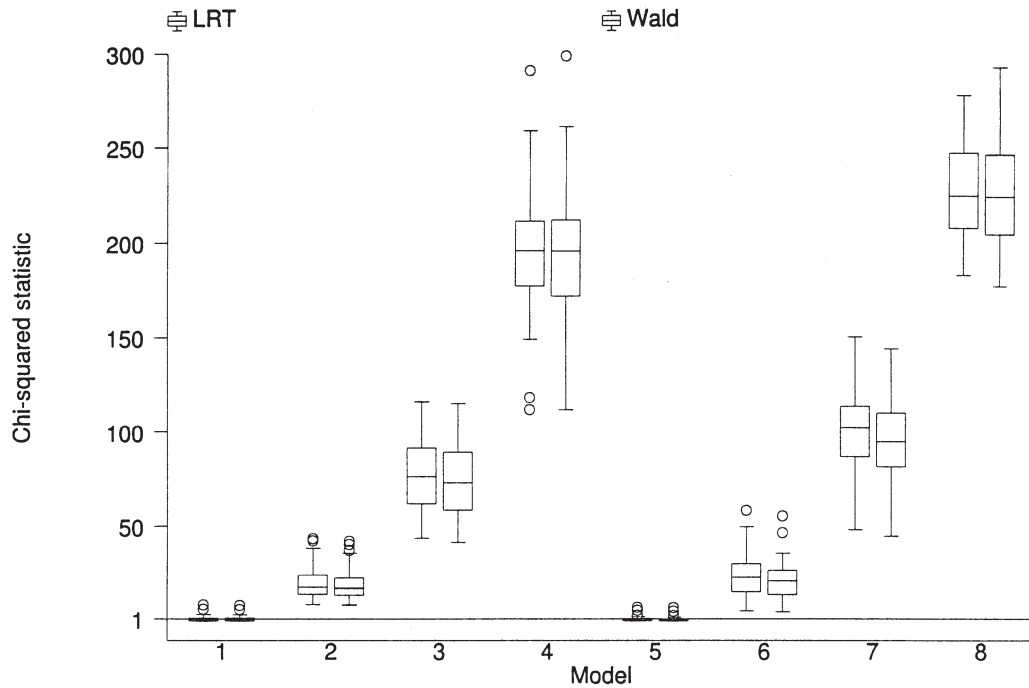
on least squares, may be more robust to violations of normality however.

**SEX-LIMITATION**

Although the regression approach can be modified to account for sex-differences in the estimates of heritability, it is not possible to straightforwardly incorporate opposite-sex DZ twins in a sex-specific analysis. Essentially, we now have five pieces of information

(the co-twin mean of each sex-zygosity) but only four non-redundant parameters ( $a^2$  and  $c^2$  for males and females) so the problem is over-identified. The model-fitting approach facilitates an easy extension to the sex-limited case including DZ-O twins, however, in a manner directly analogous to sex-limitation in standard individual differences twin model-fitting.

If the means and covariance structures for MZ and DZ same-sex male twins are coded as above, we shall introduce two new parameters  $g^2$  and  $s^2$  for additive ge-



**Fig. 1.** Basic simulation results: 10,000 pairs; for each model the left box represents the model-fitting likelihood ratio test statistic, the right box represents the regression-based Wald test statistic.

netic and shared environmental effects for females. The expected means vectors are  $[g^2 + s^2] \otimes \mathbf{P}$  for female MZ twins and  $[g^2/2 + s^2] \otimes \mathbf{P}$  for female DZ twins. For opposite-sex DZ twins, the expected means vector is  $[Zag + cs] \otimes \mathbf{P}$ , where  $Z$  is typically fixed to 0.5 although it can also be estimated to test for sex-limitation. Because the “regression to the mean” phenomenon is directly indicative of the underlying twin correlations, we are able to specify effects as path analysis would suggest for correlational data: i.e. where the DZ-O effect is the chain of paths  $Zag$ . As an example, consider the following scenario: a trait 100% heritable in males, 100% shared environmental in females. The following pattern of transformed co-twin means (i.e. reflecting the correlations) would be expected (Table IV).

A test of sex-differences in the relative balance of effects is given by comparing the model that equates  $a$  with  $g$  and  $c$  with  $s$  against a model that does not. A test of sex-specific genetic effects is given by comparing a model that fixes  $Z$  to 0.5 against a model that does not. For both female and opposite-sex twins, the residual covariance structure is the same as for males, namely

$$\begin{bmatrix} v + dp_2 & 0 \\ 0 & v + dp_1 \end{bmatrix}.$$

Appendix B gives the Mx script for sex-limited DF analysis.

If there is a mean difference by sex or zygosity (or a sex  $\times$  zygosity interaction including DZ versus DZ-O sex-specific differences, e.g. in the presence of a strong sex effect and sibling interaction) then it might be desirable to use sex- and zygosity-specific proband and population means when transforming the data. In this case, each individual is transformed using their own group-specific means, such that females, for example, regress back to the female population mean irrespective of the sex of their twin. However, arbitrary use of group-specific means is not advised, as the increasingly small group sizes will reduce the accuracy of the estimated means. Moreover, true differences can be better incorporated by

**Table IV.** Expected Co-twin Scores When  $a^2 = 1$ ,  $c^2 = 0$ ,  $e^2 = 0$  and  $g^2 = 0$ ,  $s^2 = 1$ ,  $u^2 = 0$  and  $Z = 0.5$

Sex-zygosity	Expected transformed co-twin score
MZ male	1.00
DZ male	0.50
MZ female	1.00
DZ female	1.00
DZ opposite sex	0.00

modeling the putative effect, e.g. including a sex effect or adjusting the scores prior to analysis to account for sibling interaction, based on, for example, an estimate of the sibling interaction effect from the entire sample. In an analogous manner, fitting to twin correlation matrices as opposed to covariance matrices inappropriately avoids zygosity differences in variance, rather than appropriately accounting for them. It is possible to parameterize the model in terms of the raw co-twin means, which are then transformed after analysis – this approach may facilitate incorporating group-specific effects in the model.

### SIMULATIONS: INCLUDING DZ-O TWINS

Four models were simulated 50 times each to illustrate the DF sex-limited model. In each case 2500 unselected pairs of each sex-zygosity were simulated (i.e.,  $5 \times 2500$  pairs in total) and a 5% threshold used. As shown in Table V, model 1 represents no differences between males and females; in model 2, males and females have dramatically different magnitudes of genetic effect; in model 3 males and females have less dramatic differences; in model 4 males and females have similar magnitudes of effects but completely different genes operating (i.e.  $r_{DZ-O} = 0$ ).

The results indicate accurate recovery of the sex-specific parameters. The test for different genes operating in males and females ( $LRT_1$ ) is significant at the  $p = 0.05$  level only for model 4, as predicted. The test for different magnitudes of male and female effects is significant for models 2 and 3 only, as predicted. As the critical value for a  $\chi^2$  with 2 degrees of freedom at the  $p = 0.05$  level is 5.99146, the results suggest good power to detect sex differences given this relatively large sample size. Finally, although  $r_{DZ-O}$  under model 2 is quite low (0.276), the correlation will be empirically under-identified when there is no genetic effect in males (so the concept of a correlation with the genetic effects in females is meaningless) (Table VI).

**Table V.** Four Models Simulated for Analysis under the Sex-limited DF Model

Model	$a^2$	$c^2$	$e^2$	$g^2$	$s^2$	$u^2$	$r_{DZ-O}$
1	0.50	0.00	0.50	0.50	0.00	0.50	0.5
2	0.00	0.50	0.50	0.50	0.00	0.50	0.5
3	0.50	0.25	0.25	0.25	0.50	0.25	0.5
4	0.50	0.00	0.50	0.50	0.00	0.50	0.0

Parameters for males are  $a$ ,  $c$  and  $e$ . Equivalent parameters for females are  $g$ ,  $s$  and  $u$ . The genetic correlation for DZ-O twins is  $r_{DZ-O}$ .

Table VI. Sex-Limitation Simulation Results

Model	$a^2$	$c^2$	$e^2$	$g^2$	$s^2$	$u^2$	$r_{DZ-o}$	$LRT_1$	$LRT_2$
1	0.471	0.029	0.500	0.463	0.031	0.506	0.469	0.20	1.68
2	0.054	0.459	0.488	0.490	0.010	0.500	0.276	0.11	37.64*
3	0.476	0.272	0.252	0.242	0.506	0.252	0.458	0.53	15.11*
4	0.464	0.033	0.504	0.463	0.030	0.506	0.040	8.75*	2.14

$LRT_1$  is the likelihood ratio test statistic for  $r_{DZ-o} = 0.5$  (1 df),  $LRT_2$  is the likelihood ratio test statistic for equating male ( $a^2, c^2$ ) and female ( $g^2, s^2$ ) parameter estimates,  $r_{DZ-o}$  free in both cases (2 df); tests significant at  $p = 0.05$  are marked with an asterisk (\*).

## SUMMARY

This research note presents DF analysis in a model-fitting context where an individual's expected score is a function of their co-twin's proband status. Implementations of the basic model and an extension to a sex-limited model are also presented. Other extensions that could easily be made with this implementation of the DF model include testing an ADE model instead of an ACE model, or incorporating  $G \times E$  interaction, for example.

Although we have implied a large random unselected sample from which above- (or below-) threshold probands are drawn, this need not be so. Twin pairs might only be ascertained where at least one member has an extreme score. In this case, the population mean must be estimated *a priori* in order to transform the scores appropriately. The exclusion of twin pairs concordant for nonproband status will have no impact on the estimation of  $a^2$  and  $c^2$  and the associated likelihood ratio test statistics.

In the above simulations, all genetic effects are polygenic effects homogeneous for the whole population. Under these conditions, estimates of heritability based on an unselected sample (i.e. individual differences analysis) are expected to equal estimates based on a selected extreme sample appropriately analysed (i.e., DF analysis). For real data, a significant difference between the two estimates is consistent with a number of scenarios. If, for example, a trait is found to be 'more genetic' at the extreme, this could implicate (1) different genes, (2) the same genes with different effects or (3) distributional properties of the measurement. For example, scenario (1) might describe a rare allele with a large effect, and one which is therefore present in the extreme proband group but virtually never found in the rest of the population. This genetic factor, therefore, does not account for very much variation within the entire sample, but it does within the

co-twins of probands. Scenario (2) might describe an epistatic effect: two loci  $A$  and  $B$  have main effects and epistatically interact. Extreme probands are more likely to have risk alleles at both loci, and so a greater level of genetic non-additivity is observed in the extreme group (similarly, gene-by-environment interaction could lead to this pattern if one replaces *locus B* with *environment B*). Finally, for (3), heteroscedasticity or censoring effects could lead to different estimates of heritability based on a selected part of the sample as opposed to the entire sample.

Proband status may alternatively be based on a separate dichotomous variable: for example, proband status may reflect a diagnosis of major depressive disorder whilst a continuous measure of depressive symptom severity is the co-twin trait of interest. The results would not directly estimate the variance components of the symptom severity measure, however – the relationship between severity and diagnostic status will possibly have a complex impact on parameter estimates.

Proband status could instead be defined in terms of a threshold on a different variable: for example, Purcell *et al.* (2001) illustrated the so-called *bivariate* DF model with verbal and nonverbal cognitive development measures. In the bivariate case, rather than performing two separate bivariate DF analyses (e.g. proband on  $X \rightarrow$  co-twin on  $Y$ ; proband on  $Y \rightarrow$  co-twin on  $X$ ) as well as the two univariate DF analyses (e.g. proband on  $X \rightarrow$  co-twin on  $X$ ; proband on  $Y \rightarrow$  co-twin on  $Y$ ), yielding hard-to-interpret results, a single analysis could be performed using the model-fitting method. A further advantage would be the ability to place certain testable constraints and conceptual models on the nature of the associations between proband status and continuous variation for the two variables.

In the context of the *augmented* DF model, an analogous regression-based method for unselected samples, Kohler and Rodgers (2001) note that the logic of

the Wald test's double-entry correction may be conservative and provide an alternative correction based on asymptotic regression theory. Although the model-fitting method does not utilize double-entry, some of the underlying issues raised may be relevant. In particular, it is well known that the DF model is heteroscedastic—MZ residual variance is less than DZ residual variance if the trait is heritable. Although potential heteroscedasticity may be modeled by defining the  $d$  parameter in the covariance structure

$$\begin{bmatrix} v + dp_2 & 0 \\ 0 & v + dp_1 \end{bmatrix}$$

to be zygosity-specific, this does not seem to have any significant impact on the likelihood ratio test statistics when applied to the simulated data. Additionally, residual variance components will tend to be negatively correlated between twins, whereas the current model fixes the residual twin covariance to 0. Using regression methods, the residual covariance structure can be partitioned to account for this effect; using ML methods, where means and variances are simultaneously and interdependently estimated, it is not possible to make such an adjustment. In any case, we believe the flexibility of a model-fitting implementation outweighs these concerns.

## APPENDIX A

### Basic DF Model

The following Mx script expects a transformed data file, called data.raw with six variables: ID, zygosity (1/2), transformed continuous trait scores for twin 1 and 2, proband status for twin 1 and 2. A utility to perform the transformation automatically can be downloaded from the URL <http://statgen.iop.kcl.ac.uk/df/>. Deleting the free keyword after the declaration of the A matrix (and commenting out the Matrix A 0.1 command) specifies a submodel with no genetic effect.

```
! DeFries-Fulker group analysis
! Mx Implementation
! S. Purcell, Nov 2001

G1: Matrices
Data Calc NGroups=4
Begin Matrices;
! Model parameters
A Full 1 1 free      ! additive genetic effect
C Full 1 1 free      ! shared-E effect
V Full 1 1 free      ! residual variance
```

```
D Full 1 1 free      ! variance difference
! Constants and other variables
H Full 1 1          ! constant = 1/2
I Full 1 1          ! constant = 1
N Full 1 1          ! constant = 0
Y Full 1 2          ! proband status
End Matrices;
! Starting values
Matrix A 0.1
Matrix C 0.1
Matrix V 0.8
Matrix D 0.0
! Constants
Matrix I 1
Matrix H .5
Matrix N 0
End
```

```
G2: MZ
Data NInput_vars=6 NObservations=0
Missing =-9
RE File=data.raw
Label id zyg v1 v2 p1 p2
Select if zyg = 1 /
Select v1 v2 p1 p2 /
Definition p1 p2 /
Matrices = Group 1
! DF model
Means (A+C) @ Y /
! Residual variance / covariance
Covariance
(V | N _
 N | V ) + (D | N _
             N | D ) * \v2d(Y) /
! Definition variables
Specify Y -2 -1 ! proband status
End
```

```
G3: DZ
Data NInput_vars=6 NObservations=0
Missing =-9
RE File=data.raw
Label id zyg v1 v2 p1 p2
Select if zyg = 2 /
Select v1 v2 p1 p2 /
Definition p1 p2 /
Matrices = Group 1
! DF model
Means (H*A+C) @ Y /
! Residual variance / covariance
Covariance
```



```
(V | N _
N | V ) + (D | N _
          N | D ) * \v2d(Y) /
! Definition variables
Specify Y -2 -1 ! proband status
End

G4: Collect estimates
Calculation
Matrices = group 1
Begin Algebra;
Z = A _
C _
I = (A+C) /
End Algebra;
End
```

## APPENDIX B

### Sex-Limited DF Model

The following Mx script expects a transformed data file as above. Sex-zygosity is coded 1–5 for MZ male, DZ male, MZ female, DZ female, DZ opposite-sex respectively. Deleting the free keyword after the R matrix declaration fixes the DZ-O coefficient of genetic relatedness to 0.5, to perform the test of different genes operating for males versus females. Uncommenting one or more of the Equate commands tests for a difference in magnitude of genetic and/or environmental effect.

```
! DeFries-Fulker group analysis
! Sex-limited model
! Mx Implementation
! S. Purcell, Nov 2001

G1: Matrices
Data Calc NGroups=7
Begin Matrices;
! Model parameters
A Full 1 1 free ! male additive genetic
C Full 1 1 free ! male shared-E
G Full 1 1 free ! female additive genetic
S Full 1 1 free ! female shared-E
V Full 1 1 free ! residual
D Full 1 1 free ! difference
R Full 1 1 free ! DZ OS coef. of rel.
! Constants and other variables
H Full 1 1 ! constant = 1/2
I Full 1 1 ! constant = 1
N Full 1 1 ! constant = 0
Y Full 1 2 ! proband status
```

```
End Matrices;
! Starting values and boundaries
Matrix A 0.1
Matrix C 0.1
Matrix G 0.1
Matrix S 0.1
Matrix V 0.8
Matrix D 0.0
Matrix R 0.50
Bound 0 0.5 R 1 1
! Constants
Matrix I 1
Matrix H .5
Matrix N 0
! (Un)comment to equate sex-specific estimates or not
! Equate A 1 1 G 1 1
! Equate C 1 1 S 1 1
End
```

```
G2: MZ Males
Data NInput_vars=6 NObservations=0
Missing =-9
RE File=data.raw
Label id sexzyg v1 v2 p1 p2
Select if sexzyg = 1 /
Select v1 v2 p1 p2 /
Definition p1 p2 /
Matrices = Group 1
! DF model
Means (A*A+C*C) @ Y /
! Residual variance / covariance
Covariance
(V | N _
N | V ) + (D | N _
          N | D ) * \v2d(Y) /
! Defintion variables
Specify Y -2 -1 ! proband status
End
```

```
G3: DZ Males
Data NInput_vars=6 NObservations=0
Missing =-9
RE File=data.raw
Label id zyg v1 v2 p1 p2
Select if zyg = 2 /
Select v1 v2 p1 p2 /
Definition p1 p2 /
Matrices = Group 1
! DF model
Means (H*A*A+C*C) @ Y /
```

```
! Residual variance / covariance
Covariance
(V | N _
 N | V ) + (D | N _
             N | D ) * \v2d(Y) /
! Definition variables
Specify Y -2 -1 ! proband status
End
```

```
G4: MZ Females
Data NInput_vars=6 NObservations=0
Missing =-9
RE File=data.raw
Label id sexzyg v1 v2 p1 p2
Select if sexzyg = 3 /
Select v1 v2 p1 p2 /
Definition p1 p2 /
Matrices = Group 1
! DF model
Means (G*G+S*S) @ Y /
! Residual variance / covariance
Covariance
(V | N _
 N | V ) + (D | N _
             N | D ) * \v2d(Y) /
! Definition variables
Specify Y -2 -1 ! proband status
End
```

```
G5: DZ Females
Data NInput_vars=6 NObservations=0
Missing =-9
RE File=data.raw
Label id sexzyg v1 v2 p1 p2
Select if sexzyg = 4 /
Select v1 v2 p1 p2 /
Definition p1 p2 /
Matrices = Group 1
! DF model
Means (H*G*G+S*S) @ Y /
! Residual variance / covariance
Covariance
(V | N _
 N | V ) + (D | N _
             N | D ) * \v2d(Y) /
! Definition variables
Specify Y -2 -1 ! proband status
End
```

```
G6: DZ Opposite sex
Data NInput_vars=6 NObservations=0
Missing =-9
RE File=data.raw
Label id zyg v1 v2 p1 p2
Select if zyg = 5 /
Select v1 v2 p1 p2 /
Matrices = Group 1
! DF model
Means (R*A*G+C*S) @ Y /
! Residual variance / covariance
Covariance
(V | N _
 N | V ) + (D | N _
             N | D ) * \v2d(Y) /
! Definition variables
Specify -2 -1 ! proband status
End

G7: Collect estimates
Calculation
Matrices = group 1
Begin Algebra;
Z = A*A | G*G_
C*C | S*S _
I-(A*A+C*C) | I-(G*G+S*S) _
R | R /
End Algebra;
End
```

## ACKNOWLEDGMENTS

The authors would like to acknowledge the helpful comments of John DeFries and Tom Price.

## REFERENCES

- DeFries, J. C., and Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behav. Genet.* 15(5):467-473.
- DeFries, J. C., and Fulker, D. W. (1988). Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. *Acta Genet. Med. Gemellol.* 37:205-216.
- Kohler, H. P., and Rodgers, J. L. (2001). DF-analyses of heritability with double-entry twin data: Asymptotic standard errors and efficient estimation. *Behav. Genet.* 31(2):179-191.
- Purcell, S., Eley, T. C., Dale, P. S., Oliver, B., Petrill, S. A., Price, T. S., Saudino, K. J., Simonoff, E., Stevenson, J., Taylor, E., and Plomin, R. (2001). Comorbidity between verbal and non-verbal cognitive delays in 2-year-olds: A bivariate twin analysis. *Dev. Sci.* 4:195-208.

Edited by Irwin Waldman