Acta Genet Med Gemellol 37:205-216 (1988)

Received 30 January 1988
Final 18 April 1988

Multiple Regression Analysis of Twin Data: Etiology of Deviant Scores versus Individual Differences

J.C. DeFries, D.W. Fulker

Institute for Behavioral Genetics, University of Colorado, Boulder, USA

Abstract. The multiple regression analysis of twin data in which a cotwin's score is predicted from a proband's score and the coefficient of relationship (the basic model) provides a statistically powerful test of genetic etiology. When an augmented model that also contains an interaction term is fitted to the same data set, direct estimates of heritability (h^2) and the proportion of variance due to shared environmental influences (c^2) are obtained. A simple transformation of selected twin data prior to regression analysis facilitates direct estimates of h_g^2 (an index of the extent to which the difference between the mean of probands and that of the unselected population is heritable) and a test of the hypothesis that the etiology of deviant scores differs from that of variation within the normal range.

Key words: Genetic etiology, Heritability, Individual differences, Multiple regression, Twins

INTRODUCTION

DeFries and Fulker [3] recently proposed a multiple regression analysis of selected twin data in which a cotwin's score is predicted from that of a proband (the member of the pair selected because of a deviant score on a continuous variable) and the coefficient of relationship. Two models were formulated: (1) a basic model in which the partial regression of cotwin's score on the coefficient of relationship provides a test for genetic etiology; and (2) an augmented model containing an interaction term between proband's score and the coefficient of relationship that yields direct estimates of heritability (h^2) and the proportion of variance due to

This work was supported in part by a program project grant from the NICHD (HD-11681).

environmental influences shared by members of twin pairs (c^2) potentially relevant to the unselected population.

DeFries and Fulker [3] also illustrated how the results of fitting the basic model to selected twin data could be used to obtain an estimate of the extent to which the deficit exhibited by the probands is due to heritable influences (h_g^2). Moreover, it was suggested that a comparison of h_g^2 and h^2 could be used to test the hypothesis that the etiology of extreme scores may differ from that of variation within the normal range. Markedly low test scores, for example, could be caused by a major gene, a chromosomal anomaly, or a special environmental insult, whereas individual differences within the normal range might be due to multifactorial influences. On the other hand, if probands merely represent the lower tail of a normal distribution of individual differences, h_g^2 and h^2 should be similar in magnitude. Although DeFries and Fulker [3] advocated a comparison of h_g^2 and h^2, no statistical test of this difference was proposed.

Following publication of the initial report, we employed this methodology to analyze data from twin pairs in which at least one member of each pair is reading disabled [4]. During further investigation, it became apparent that a simple transformation of twin data prior to multiple regression analysis facilitates direct estimates of h_g^2 and a test of the difference between h_g^2 and h^2. Thus, the primary objectives of the present report are twofold: (1) to illustrate the application of this transformation to simulated twin data; and (2) to assess the power of the multiple regression analysis of selected twin data to detect significant h_g^2, h^2, and their difference.

MODELS

When probands have been selected because of deviant scores on a continuous variable, the differential regression of identical (MZ) and fraternal (DZ) cotwin's scores toward the mean of the unselected population provides a test for genetic etiology. For example, when probands have been selected on the basis of low test scores, DZ cotwins are expected to have higher scores than MZ cotwins to the extent that the condition is heritable. Thus, a t-test of the difference between the means for MZ and DZ cotwins would suffice as a test for genetic etiology when the probands have identical mean scores. However, the partial regression of cotwin's score on the coefficient of relationship ($R = 1.0$ and 0.5 for MZ and DZ twin pairs, respectively), independent of proband's score, provides a more general and powerful test. Moreover, when the product of the proband's score and the coefficient of relationship is added to the model during a second step in the analysis, direct estimates of h^2 and c^2 are also obtained. Regression analyses of such attenuated data are more appropriate than correlation analyses because regression coefficients are less influenced by restriction of range of the independent variables [2: p 65, 7: p 60].

The basic model in which a cotwin's score (C) is predicted from the proband's score (P) and the coefficient of relationship (R) is as follows:

$$C = B_1 P + B_2 R + A\,, \tag{1}$$

where B_1 is the partial regression of cotwin's score on proband's score, B_2 is the partial regression of cotwin's score on the coefficient of relationship, and A is the regression constant.

In contrast to the basic model, the augmented model contains an interaction term:

$$C = B_3 P + B_4 R + B_5 PR + A\,, \tag{2}$$

where PR is the product of proband's score and the coefficient of relationship. As noted below, inclusion of the interaction term in the augmented model changes the expectations for the partial regression coefficients estimated from the basic model. Thus, the coefficients of P and R are symbolized B_3 and B_4 in Equation (2).

DeFries and Fulker [3] asserted that B_1 is a measure of twin resemblance that is independent of zygosity and that B_2 equals twice the difference between the means for MZ and DZ cotwins after covariance adjustment for any difference that may exist between MZ and DZ probands. Thus, B_2 was advocated as a test of significance for genetic etiology analogous to that of differential twin concordance. In addition, it was suggested that the ratio of B_2 to the difference between the mean for probands and that for the unselected population estimates h_g^2. The authors also illustrated that B_3 and B_5 provide direct estimates of c^2 and h^2.

LaBuda et al [6] subsequently derived the expected partial regression coefficients for the basic and augmented models as functions of additive genetic variance (V_A), variance due to environmental influences shared by members of twin pairs (V_C), phenotypic variance (V_P), and the cotwin and proband means. The expected partial regression coefficients estimated from the fit of the basic model to selected twin data are as follows:

$$B_1 = [(n_1 + n_2/2)/N]\, V_A/V_P + V_C/V_P \tag{3}$$

$$B_2 = 2\left[(\overline{C}_{MZ} - \overline{C}_{DZ}) - B_1(\overline{P}_{MZ} - \overline{P}_{DZ})\right]\,, \tag{4}$$

where n_1 is the number of MZ twin pairs, n_2 is the number of DZ pairs, $N = n_1 + n_2$, $\overline{C}_{MZ}$ and $\overline{C}_{DZ}$ are means for the MZ and DZ cotwins, and $\overline{P}_{MZ}$ and $\overline{P}_{DZ}$ are means for MZ and DZ probands. Given the standard assumptions of quantitative genetic analyses of twin data (eg, a linear polygenic model, little or no assortative mating, and equal shared environmental influences for MZ and DZ twin pairs), the expected MZ covariance contains $V_A + V_C$, whereas that for DZ pairs is $\frac{1}{2}V_A + V_C$. Thus, B_1 is a weighted average of twin resemblance in the combined sample of MZ and DZ twin pairs. B_2, on the other hand, equals twice the difference between the means for MZ and DZ cotwins after covariance adjustment for the difference between the means for MZ and DZ probands. As indicated below, B_2 provides a direct estimate of h_g^2 when the data are suitably transformed.

The expected partial regression coefficients estimated from the fit of the augmented model are as follows:

$$B_3 = V_C/V_P = c^2 \tag{5}$$

$$B_4 = 2\{(\overline{C}_{MZ} - \overline{C}_{DZ}) - [\overline{P}_{MZ}(h^2 + c^2) - \overline{P}_{DZ}(h^2/2 + c^2)]\} \quad (6)$$

$$B_5 = V_A/V_P = h^2 \quad (7)$$

Thus, as LaBuda et al [6] noted, B_3 and B_5 provide unbiased estimates of c^2 and h^2, respectively. b_4, similar to B_2, is a function of twice the difference between the means for MZ and DZ cotwins. However, unlike B_2, it does not provide a test for genetic etiology. In fact, when the twin data are suitably transformed, B_4 tests the difference between h_g^2 and h^2.

DATA TRANSFORMATION

When untransformed data are analyzed, the expected means of MZ and DZ cotwins can be predicted from observed proband means as follows:

$$C_{MZ} = \mu + (h_g^2 + c_g^2)(\overline{P}_{MZ} - \mu) \quad (8)$$

$$C_{DZ} = \mu + (\tfrac{1}{2}h_g^2 + c_g^2)(\overline{P}_{DZ} - \mu), \quad (9)$$

where c_g^2, analogous to h_g^2, is an index of the extent to which the difference between the mean for probands and that of the unselected population (μ) is due to environmental influences shared by members of twin pairs.
Upon rearrangement,

$$h_g^2 + c_g^2 = (\overline{C}_{MZ} - \mu)/(\overline{P}_{MZ} - \mu) \quad (10)$$

$$\tfrac{1}{2}h_g^2 + c_g^2 = (\overline{C}_{DZ} - \mu)/(\overline{P}_{DZ} - \mu). \quad (11)$$

Thus, in general,

$$h_g^2 = 2\left[(\overline{C}_{MZ} - \mu)/(\overline{P}_{MZ} - \mu) - (\overline{C}_{DZ} - \mu)/(\overline{P}_{DZ} - \mu)\right]. \quad (12)$$

However, when $\overline{P}_{MZ} = \overline{P}_{DZ} = \overline{P}$,

$$h_g^2 = 2\left[(\overline{C}_{MZ} - \overline{C}_{DZ})/(\overline{P} - \mu)\right] \quad (13)$$

For the sake of simplicity, we assume that MZ and DZ probands have equal means and variances in the following derivations. For most applications, this assumption should be correct within sampling error. However, as noted below, the analysis is sufficiently general to accommodate differences between MZ and DZ proband means or variances when they occur.

From Equations (4) and (13) it follows that

$$B_2 = 2(\overline{C}_{MZ} - \overline{C}_{DZ}) = h_g^2(\overline{P} - \mu) \quad (14)$$

when $\overline{P}_{MZ} = \overline{P}_{DZ} = \overline{P}$. Thus, as previously noted, the ratio of B_2 to the difference between the mean for probands and that of the unselected population yields an estimate of h_g^2. Alternatively, when the scores of probands and cotwins are each divided by $\overline{P} - \mu$ prior to fitting the basic model, B_2 directly estimates h_g^2.

As already indicated, the expectation for B_4 is not as simple as that for B_2. However, as may be seen from Equation (6),

$$B_4 = 2\left[(\overline{C}_{MZ} - \overline{C}_{DZ}) - \frac{1}{2}h^2(\overline{P})\right] \tag{15}$$

when $\overline{P}_{MZ} = \overline{P}_{DZ}$. By expressing the difference between the MZ and DZ cotwin means as a function of h_g^2 and $\overline{P} - \mu$,

$$B_4 = 2\left[(\frac{1}{2}h_g^2(\overline{P} - \mu) - \frac{1}{2}h^2(\overline{P})\right] = h_g^2(\overline{P} - \mu) - h^2(\overline{P}). \tag{16}$$

Thus, B_4 is a function of $\overline{P}$ and μ with untransformed data, ie, it is dependent upon an arbitrary scale of measurement.

By transforming the proband and cotwin data, the expectation for B_4 becomes more meaningful. When scores are expressed as a deviation from the mean of the unselected population, $\mu = 0$. Thus, when the augmented model is fitted to data transformed in this manner,

$$B_4 = h_g^2(\overline{P}) - h^2(\overline{P}) = (h_g^2 - h^2)\overline{P}. \tag{17}$$

Consequently, when the scores of MZ and DZ twins are each expressed as a deviation from the mean of the unselected population, the significance of B_4 provides a test of the difference between h_g^2 and h^2.

From Equation (17) it may also be seen that B_4 will estimate $h_g^2 - h^2$ directly if each score is expressed as a deviation from the unselected population mean and then divided by $\overline{P}$. Obviously, when each score is expressed as a deviation from μ and divided by a constant, the expectations for B_1, B_3, and B_5 are unchanged. Thus, fitting the basic and augmented models to MZ and DZ twin data transformed in this very simple manner facilitates direct estimates as h_g^2, c^2, h^2, and $h_g^2 - h^2$.

APPLICATION

To illustrate the application of this transformation, the basic and augmented models are fitted to three data sets: (1) a set of untransformed data in which the mean of the unselected population is assumed to be 16; (2) the same data expressed as deviations from $\mu = 16$; and (3) the same scores expressed as deviations from μ and then divided by $\overline{P} - \mu$.

Table 1 - Simulated raw and transformed scores to illustrate multiple regression analysis of twin data

	Cotwin	Proband	Coefficient of relationship
Raw data			
	13.00	10.0	1.0
	8.60	8.0	1.0
	10.80	6.0	1.0
	4.20	4.0	1.0
	6.40	2.0	1.0
	9.60	10.0	0.5
	12.10	8.0	0.5
	14.70	6.0	0.5
	4.50	4.0	0.5
	7.10	2.0	0.5
Expressed as a deviation from $\mu = 16$			
	-3.00	-6.0	1.0
	-7.40	-8.0	1.0
	-5.20	-10.0	1.0
	-11.80	-12.0	1.0
	-9.60	-14.0	1.0
	-6.40	-6.0	0.5
	-3.90	-8.0	0.5
	-1.30	-10.0	0.5
	-11.50	-12.0	0.5
	-8.90	-14.0	0.5
Expressed as a deviation from μ and divided by $\overline{P} - \mu = -10$			
	0.30	0.6	1.0
	0.74	0.8	1.0
	0.52	1.0	1.0
	1.18	1.2	1.0
	0.96	1.4	1.0
	0.64	0.6	0.5
	0.39	0.8	0.5
	0.13	1.0	0.5
	1.15	1.2	0.5
	0.89	1.4	0.5

First, consider the simulated raw data listed at the top of Table 1 and the corresponding summary statistics presented in Table 2. Note that the MZ and DZ proband means are each 10 units below that of the unselected population. As expected for a heritable disorder, the regression of the MZ cotwin mean toward the mean of the unselected population is less than that for DZ cotwins, viz., 2.6 vs. 3.6, respectively. Thus, when the basic model shown in Equation (1) is fitted to these simulated data, $B_2 = 2(\overline{C}_{MZ} - \overline{C}_{DZ}) = -2.0$. Dividing B_2 by $\overline{P} - \mu = -10$ yields an estimate for h_g^2 of 0.2. This result suggests that about 20% of the difference

between the proband mean and that of the unselected population is due to heritable influences.

Table 2 - Summary statistics from simulated twin data[a]

	Zygosity	$\overline{P}$	$\overline{C}$	s_P^2	s_C^2	b_{CP}
Raw data						
	MZ	6	8.6	10.0	12.100	0.88
	DZ	6	9.6	10.0	16.130	0.63
Expressed as a deviation from $\mu = 16$						
	MZ	-10	-7.4	10.0	12.100	0.88
	DZ	-10	-6.4	10.0	16.130	0.63
Expressed as a deviation from μ and divided by $\overline{P} - \mu = -10$						
	MZ	1	0.74	0.1	0.121	0.88
	DZ	1	0.64	0.1	0.161	0.63

[a] $\overline{P}$ is the proband mean, $\overline{C}$ is the cotwin mean, s_P^2 is the proband variance, s_C^2 is the cotwin variance, and b_{CP} is the regression of cotwin's score on proband's score.

From Table 2 it may also be seen that the regression of cotwin's score on proband's score for MZ twins is 0.88, whereas that for DZ twins is 0.63. Because the numbers of MZ and DZ twin pairs are equal in this simulated data set, the estimate of $B_1 = 0.755$ is the arithmetic average of these two regression coefficients.

Doubling the difference between the MZ and DZ regression coefficients estimates $h^2 = 0.50$ and subtracting this value from the MZ regression estimates $c^2 = 0.38$. As expected, when the augmented model [Equation (2)] is fitted to these data, $B_5 = h^2 = 0.50$ and $B_3 = c^2 = 0.38$. From Equation (16), $B_4 = h_g^2(\overline{P} - \mu) - h^2(\overline{P}) = 0.2(6-16) - 0.5(6) = -5.0$, a parameter estimate of relatively little interest. However, as shown below, the estimate for B_4 is of greater interest when the data are suitably transformed.

The scores of MZ and DZ probands and cotwins are each expressed as a deviation from $\mu = 16$ in the middle portion of Table 1, and the corresponding summary statistics are presented in Table 2. As expected, given this simple linear transformation, neither the variances nor the regression coefficients change. Thus, when the basic and augmented models are fitted to these transformed data, the estimates for B_1, B_2, B_3, and B_5 also do not change. However, from Equation (17), $B_4 = (h_g^2 - h^2)\overline{P} = (0.2 - 0.5)(-10) = 3.0$ when estimated from this transformed data set. Since the null hypothesis is that $h_g^2 = h^2, B_4$ may be used as a test of significance for the difference between these two parameters. However, dividing each deviation score by $\overline{P} - \mu$ results in an even simpler interpretation.

The deviation scores each divided by $\overline{P} - \mu$ are listed in the bottom portion of Table 1, and the corresponding summary statistics are presented in Table 2. As

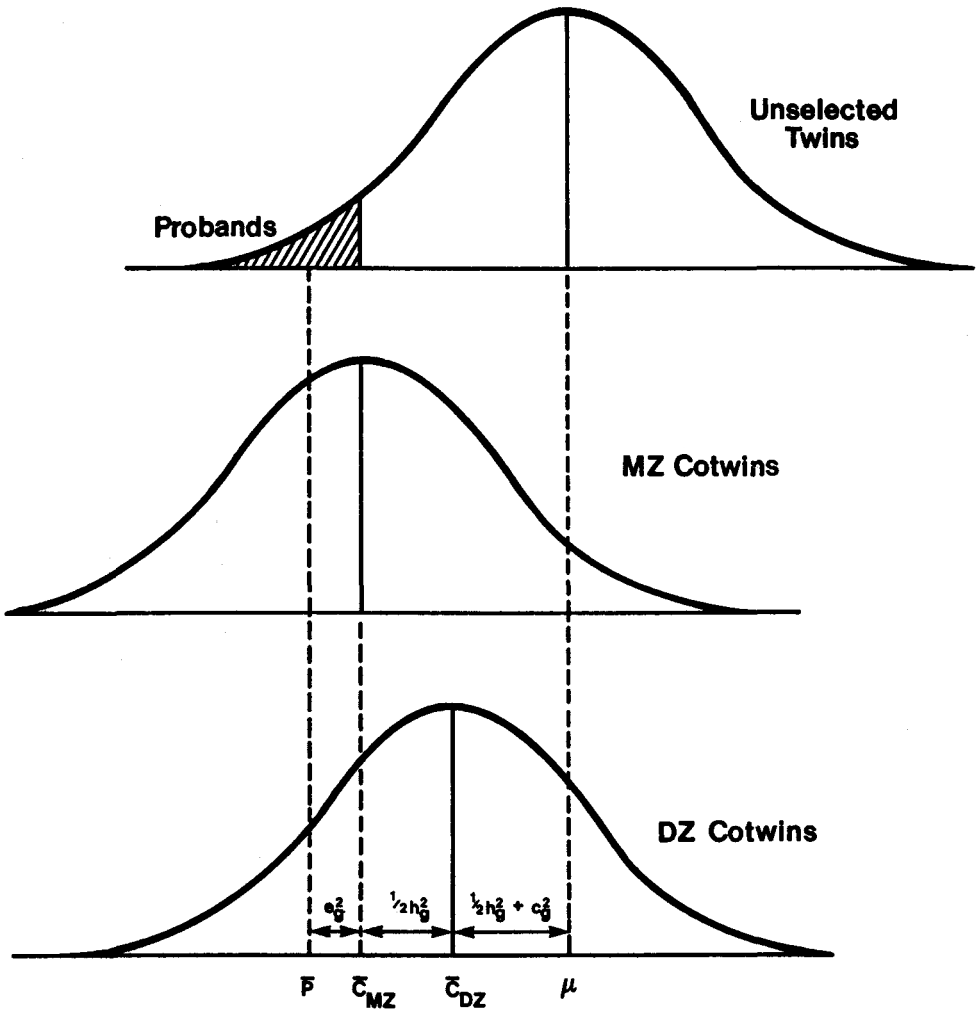

Figure. Hypothetical distributions of transformed data from an unselected population of twins (with mean μ) and from the monozygotic (MZ) and dizygotic (DZ) cotwins of proband. Proband and cotwin means are symbolized $\bar{P}$ and $\bar{C}$, respectively. The deficit of probands ($\bar{P} - \mu$) is due to heritable influences (h_g^2) and to environmental influences that are either shared (c_g^2) or not shared (e_g^2) by members of twin pairs.

expected, given the nature of this transformation, the variances and covariances involving P and C in the third data set are the original values, each divided by $(\overline{P} - \mu)^2 = 100$. However, because these changes are proportional, the MZ and DZ regressions remain unchanged, as do the estimates for B_1, B_3, and B_5. In contrast, B_2 and B_4 are functions of means and therefore change with this transformation.

As expected, the transformed $\overline{P}$ is unity. Thus, from Equation (14), B_2 provides a direct estimate of $h_g^2 = 0.2$ when this transformation is employed. For the same reason, it may be seen from Equation (17) that B_4 now estimates $h_g^2 - h^2 = -0.3$.

By employing this simple transformation, the MZ and DZ cotwin means can be used to partition the "standadized deviation" of the proband mean [ie, $\overline{P} - \mu = 1$] into its component parts. From Equations (10) and (11) it follows that $\overline{C}_{MZ} = h_g^2 + c_g^2$ and that $\overline{C}_{DZ} = \frac{1}{2}h_g^2 + c^2$ when the data are transformed in this manner. These relationships are illustrated in the Figure, where it also may be seen that $\overline{P} - C_{MZ} = 1.0 - (h_g^2 + c_g^2) = e_g^2$, an index of the extent to which $\overline{P} - \mu$ is due to environmental influences not shared by members of twin pairs. Obviously, this partitioning of $\overline{P} - \mu$ into components due to heritable and environmental influences is exactly analogous to the partitioning of phenotypic variance using MZ and DZ intraclass correlations.

Throughout, we have assumed that the MZ and DZ proband means are equal. For most applications, this assumption will be approximately correct. However, even if the MZ and DZ probands were selected differentially, and therefore had different means, a multiple regression analysis of transformed data could still be undertaken. In this case, the MZ proband and cotwin scores should each be divided by $\overline{P}_{MZ} - \mu$, whereas the DZ proband and cotwin scores should each be divided by $\overline{P}_{DZ} - \mu$. By fitting the basic and augmented models to twin data transformed in this manner, direct estimates of $B_2 = h_g^2, B_3 = c^2, B_5 = h^2$, and $B_4 = h_g^2 - h^2$ would be obtained. However, because the variances of the transformed MZ and DZ proband scores are rendered unequal by the differential transformation, the estimate for B_1 is no longer the arithmetic average of the MZ and DZ regression coefficients.

POWER

Experience to date with applying the multiple regression analysis to reading performance data from reading-disabled probands and their cotwins suggests that B_2 is a powerful test of genetic etiology. This general impression is substantiated by a comparison of the statistical power [1] of alternative tests, where power equals the probability of rejecting the null hypothesis, ie, 1 - type II error. First, consider the power to detect a significant difference in concordance rates between MZ and DZ twin pairs. When discriminant function weights estimated from reading performance data on an independent sample of affected and control singletons are used to diagnose reading disability, the concordance rates in a sample of 62 MZ and 55 DZ

twin pairs are about 50% and 35%, respectively [5]. Given these proportions, the power to detect a significant difference at the 0.05 level (one-tailed test because MZ concordance must exceed DZ concordance for evidence of a genetic etiology) with a sample of 100 pairs of MZ and 100 pairs of DZ twins is 0.68.

An alternative test of genetic etiology is based upon a simple comparison of MZ and DZ cotwin means. To the extent that reading disability is heritable, the mean for the DZ cotwins should regress more than that for the MZ cotwins toward the mean of the unselected population. Thus, given that the MZ and DZ proband means are approximately equal, a t-test of the difference between the MZ and DZ cotwin means also provides a test for genetic etiology. In our sample of reading-disabled probands and cotwins, the average discriminant score of DZ cotwins exceeds that of MZ cotwins by about 0.35 standard deviation units. Given this difference, the power to detect significance at the 0.05 level (one-tailed test) in a sample of 100 pairs of MZ and 100 pairs of DZ twins is 0.78.

Finally, when the basic model is fitted to the transformed discriminant function score data from this sample of reading-disabled probands and cotwins, the estimate for $B_2 = h_g^2$ is about 0.30. Given an observed squared multiple correlation of 0.44 and a correlation between proband and cotwin scores of 0.63, the power to detect a significant B_2 at the 0.05 level (one-tailed test) in a sample of 100 pairs of MZ and 100 pairs of DZ twins is 0.96. Thus, the B_2 estimate from such a sample would have a type II error of only 0.04, whereas the type II error associated with a t-test of the difference between the MZ and DZ cotwin means in the same sample would be 0.22. The multiple regression analysis of twin data has much greater power to detect significance because proband scores are employed as a covariate and account for a substantial proportion of the variance in cotwin scores when the twin correlation is moderate to large. This variance in cotwin scores is included in the error variance in the t-test of the difference between the MZ and DZ cotwin means, but is partialed out in the multiple regression analysis.

Although the multiple regression analysis of selected twin data provides a powerful test of genetic etiology, the probability of rejecting the null hypothesis that $h^2 = 0$ using this approach is no greater than that for estimates obtained from alternative twin analyses. For example, given that $h^2 = 0.5$ and the squared multiple correlations estimated from the data in Table 1, the power to detect a significant $B_5 = h^2$ at the 0.05 level (one-tailed test) in a sample of 100 pairs of MZ and 100 pairs of DZ twins is only 0.45. In a sample of 200 pairs of MZ and 200 pairs of DZ twins, the power increases to a more respectable 0.73.

Because the power to detect a significant h^2 is relatively low, we can anticipate that the power to detect a significant difference between h_g^2 and h^2 will be even lower. Given the data in Table 1 in which $h_g^2 = 0.2$ and $h^2 = 0.5$ and the associated squared multiple correlations, the power to detect a significant $B_4 = h_g^2 - h^2$ at the 0.05 level (two-tailed because there is no a priori expectation about the direction of the difference) in a sample of 200 pairs of MZ and 200 pairs of DZ twins is 0.23. With 500 pairs of each zygosity type, the power increases only to 0.54. Thus, very large samples would be required to test the hypothesis that probands merely represent the lower tail of a normal distribution of individual differences.

SINGLE OR DOUBLE ENTRY

For the sake of simplicity, the multiple regression analyses of selected twin data employed throughout the present study are "single entry", ie, a score for each member of each twin pair is entered only once, either as a proband or as a cotwin. This procedure is appropriate when only one proband is ascertained per twin pair. If both members of the pair are ascertained because of deviant scores, their data could be entered twice, once as a proband and once as a cotwin. However, this "double entry" detracts somewhat from the elegant simplicity of the multiple regression analysis of selected twin data and our experience to date, although limited, suggests that parameters estimated from these two procedures may not differ greatly [4].

DISCUSSION

The multiple regression analysis of selected twin data provides a powerful and flexible test of genetic etiology. When probands have been selected because of deviant scores on a continuous variable, MZ and DZ cotwins are expected to regress differentially toward the mean of the unselected population to the extent that the condition is heritable. Fitting a regression model to such data, in which a cotwin's score (C) is predicted from a proband's score (P) and the coefficient of relationship (R), the partial regression of C on R estimates twice the difference between the means for MZ and DZ cotwins after covariance adjustment for any difference between MZ and DZ probands. Thus, this partial regression coefficient provides a direct test of genetic etiology.

The partial regression of C on R is also equivalent to $h_g^2(\overline{P} - \mu)$, where h_g^2 is an index of the extent to which the difference between the proband mean and that of the unselected population (ie, $\overline{P} - \mu$) is due to heritable influences. When the scores of probands and cotwins are each divided by $\overline{P} - \mu$ prior to fitting the basic model, the partial regression of C on R thus estimates h_g^2 directly.

The basic model can be extended in a number of different ways to facilitate more comprehensive analyses of twin data. For example, age adjustment of data could be easily accomplished by including age of proband as another independent variable in the regression model. By also including the product of age and the coefficient of relationship in the model, it would be possible to test for differential genetic etiology as a function of age. Such a model could be employed to test the recent hypothesis of Stevenson et al [8] that genetic factors may be less influential as a cause of reading disability in 13-year-old children than in children at younger ages.

The basic model could also be readily extended to include main effects and interactions involving gender, socioeconomic status, ethnic group, etc. A possible variable of special relevance to the genetics of reading disability is subtype. If data from probands of ostensibly different subtype were analyzed simultaneously, the

partial regression of cotwin's score on the product of R and a dummy variable representing subtype would provide a test for differential genetic etiology. A positive finding from such an analysis would provide compelling evidence for subtype validity. The power to detect significant differential genetic etiology, of course, depends upon the magnitude of h_g^2 in the different subtype. For example, if h_g^2 differed by 0.5 in two subtype, the power to detect a significant interaction between R and subtype at the 0.05 level (two-tailed test) in a sample of 100 pairs of MZ and 100 pairs of DZ twins would be about 0.75. However, if the difference were 0.3, the power would be only about 0.30. By increasing the sample size to 150 pairs in each zygosity group, the power is increased to about 0.90 and 0.50 in these two cases.

By fitting an augmented model containing an interaction term between proband's score and the coefficient of relationship to twin data, direct estimates of h^2 and c^2 can also be obtained. We previously noted that a comparison of h_g^2 and h^2 could be used to test the hypothesis that the etiology of extreme scores differs from that of variation within the normal range. When the data are transformed by expressing each score as a deviation from the mean of the unselected population and dividing this deviation by $\overline{P} - \mu$, the partial regression of C on R from the augmented model directly estimates $h_g^2 - h^2$. Because the power to detect a significant difference between $h_g^2 - h^2$ is relatively low, data from a large sample of probands and cotwins would be required to test the hypothesis. Nevertheless, the test is of considerable theoretical interest and may be feasible for twin studies of relatively common disorders.

Acknowledgments. We thank Dianne Johnson for typing the manuscript and Rebecca G. Miles for expert editorial assistance.

REFERENCES

1. Cohen J (1977): Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press.
2. Cohen J,. Cohen P (1975): Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.
3. DeFries JC, Fulker DW (1985): Multiple regression analysis of twin data. Behav Genet 15:467-473.
4. DeFries JC, Fulker DW, LaBuda MC (1987): Evidence for a genetic aetiology in reading disability of twins. Nature 329:537-539.
5. LaBuda MC, DeFries JC (in press): Genetic etiology of reading disability: Evidence from a twin study. In Pavlidas GTh (ed): Dyslexia. New York: Wiley.
6. LaBuda MC, DeFries JC, Fulker, DW (1986): Multiple regression analysis of twin data obtained from selected samples. Genet Epidemiol 3:425-433.
7. Morton NE (1982): Outline of Genetic Epidemiology. New Yoek: Karger.
8. Stevenson J, Graham P, Fredman G, McLoughlin V (1987): A twin study of genetic influences on reading and spelling ability and disability. J Child Psychol Psychiat 28:229-247.

Correspondence: Prof. J.C. DeFries, Institute for Behavioral Genetics, University of Colorado, Boulder, 80309-0447, USA