www.AUSTRALIANACADEMICPRESS.com.au

ARTICLE AVAILABLE ONLINE
Twin Research and Human Genetics
Volume 15 ▮ Number 1 ▮ pp. 116–119

Multiple Regression Analysis of Reading Performance Data from Twin Pairs with Reading Difficulties and Nontwin Siblings: The Augmented Model

Sally J. Wadsworth[1], Richard K. Olson[1, 2], Erik G. Willcutt[1, 2] and John C. DeFries[1, 2]
[1] *Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA*
[2] *Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA*

The augmented multiple regression model for the analysis of data from selected twin pairs was extended to facilitate analyses of data from twin pairs and nontwin siblings. Fitting this extended model to data from both selected twin pairs and siblings yields direct estimates of heritability (h^2) and the difference between environmental influences shared by members of twin pairs and those of sib or twin–sib pairs (i.e., $c^2_{(t)} - c^2_{(s)}$). When this model was fitted to reading performance data from 293 monozygotic and 436 dizygotic pairs selected for reading difficulties, and 291 of their nontwin siblings, $h^2 = .48 \pm .22$, $p = .03$, and $c^2_{(t)} - c^2_{(s)} = .22 \pm .12$, $p = .06$. Although the test for differential shared environmental influences is only marginally significant, the results of this analysis suggest that environmental influences on reading performance that are shared by members of twin pairs (.36) may be substantially greater than those for less contemporaneous twin–sibling pairs (.14).

■ **Keywords:** multiple regression, twins, siblings, reading, heritability, shared environment

The basic and augmented multiple regression models that DeFries and Fulker (1985; 1988) advocated for analyzing data from selected twin pairs have become standards of behavioral genetic analysis. The basic model is appropriate for assessing the heritable nature of group deficits, whereas the augmented model is used to assess heritability and shared 'environmentality' of individual differences within the selected group. The basic and augmented models are as follows:

$$C = B_1P + B_2R + A \tag{1}$$

$$C = B_3P + B_4R + B_5PR + A \tag{2}$$

where C is the cotwin's score, P is the proband's score, R is the coefficient of relationship (1.0 and .5 for identical [MZ] and fraternal [DZ] twin pairs, respectively), and A is the regression constant. From the basic model, when the data are transformed by expressing each score as a deviation from the unselected population mean and dividing by the difference between the proband mean and the mean of the unselected population (i.e., the selection differential), $B_2 = h^2_g$, an estimate of the extent to which the deviant scores of probands are due to heritable influences. From the augmented model, $B_5 = h^2$, a measure of the extent to which individual differences within the selected group are heritable, $B_3 = c^2_{(t)}$, a corresponding estimate of the extent to which environmental influences are shared by members of twin pairs, and $B_4 = h^2_g - h^2$.

Recently, Astrom et al. (2011, 2012) extended the basic model to facilitate analyses of data from selected twin pairs and nontwin siblings as follows:

$$C = B_1P + B_2R + B_6S + A \tag{3}$$

where C is the cotwin's or cosib's score, P is the proband's score, R is the coefficient of relationship (1.0 for MZ pairs, and .5 for DZ pairs and sib or twin–sib pairs), and S is a

RECEIVED 17 June, 2011; ACCEPTED 03 November, 2011.

ADDRESS FOR CORRESPONDENCE: Sally J. Wadsworth, Institute for Behavioral Genetics, 447 UCB, University of Colorado, Boulder CO 80309, USA. E-mail: Sally.Wadsworth@Colorado.edu

Wadsworth, S.J., Olson, R.K., Willcutt, E.G. & DeFries, J.C. (2012). Multiple regression analysis of reading performance data from twin pairs with reading difficulties and nontwin siblings: The augmented model. *Twin Research and Human Genetics*, *15*, 1, 116–119. DOI 10.1375/twin.15.1.116

dummy code for pair type (+.5 for both MZ twin and DZ twin pairs, and –.5 for sib or twin–sib pairs). When this model is fitted to transformed twin–sib data, B_2 estimates h^2_g from the twin data (i.e., twice the difference between the MZ and DZ cotwin means), and B_6 estimates the difference between the DZ cotwin and cosib means, thereby providing a test of the difference between environmental influences shared by members of DZ twin pairs and those of sib or twin–sib pairs (i.e., $c^2_{g(t)} - c^2_{g(s)}$).

For example, Astrom et al. (in press) recently analyzed discriminant function data (a composite measure of reading performance) from 245 MZ and 420 DZ twin pairs in which at least one twin met proband criteria, and from 303 of their nontwin siblings. Results obtained from analyses of the twin-only data yielded estimates of h^2_g = .667 and $c^2_{g(t)}$ = .252. When Equation 3 was fitted to the combined twin and cosib-twin data, $B_2 = h^2_g = .667$ and $B_6 = c^2_{g(t)} - c^2_{g(s)} = .082$. Although the B_6 coefficient was relatively small, it was significant (p = .02), suggesting that environmental influences shared by members of twin pairs (.252) are somewhat greater than those for less contemporaneous twin/nontwin sibling pairs (.252 – .082 = .170) for this measure.

In a corresponding manner, Equation 2 may also be easily extended to assess the etiologies of individual differences in selected twin pairs and nontwin siblings. Thus, the primary objective of this brief note is to present this extended augmented model and illustrate its application.

Materials and Methods

Subjects and Measures

The subjects participated in either the Colorado Reading Project (DeFries, Olson, Pennington, & Smith, 1991) or the ongoing Colorado Learning Disabilities Research Center (DeFries et al., 1997). In order to minimize the possibility of ascertainment bias, the sample was systematically obtained through 27 school districts in the state of Colorado. School administrators identified twin pairs in each school without regard to reading status, and then permission was sought from parents to review the school records of their children for evidence of reading problems (e.g., low reading achievement test scores or referral to resource rooms because of low reading performance). Parents were also asked whether either twin had difficulties learning to read. Twin pairs in which at least one member had a history of reading problems were invited to complete extensive test batteries of reading-related and cognitive measures in laboratories at both the University of Colorado, Boulder, and the University of Denver. Data from the Peabody Individual Achievement Test (Dunn & Markwardt, 1970) Reading Recognition, Reading Comprehension and Spelling subtests were then used to compute a discriminant function score for each child, employing weights estimated from an analysis of data from an independent sample tested during an earlier phase of the project (DeFries, 1985). Twin pairs with a positive history of reading problems were included in the proband sample if at least one member of the pair was also classified as affected by the discriminant score and had a Verbal or Performance IQ of at least 85, no serious neurological, emotional, or behavioral problems, and no uncorrected visual or auditory acuity deficits. Control twins were matched to probands on the basis of age, gender, and school district, and had a negative history for reading problems.

Selected items from the Nichols and Bilbro (1966) twin questionnaire were used to classify zygosity. In doubtful cases, zygosity was confirmed by genotyping blood or buccal samples. Twin pairs were reared in English-speaking, primarily middle-class homes, and ranged in age from 8 to 20 years. Prior to testing, the session was fully described to family members, and informed consent was obtained from subjects aged 18 years or older. For children who were younger than 18 years, assent was obtained from the children and permission was obtained from their parents. The resulting sample included 293 MZ and 436 DZ pairs in which at least one member of each pair met proband criteria, and 291 of their nontwin siblings. In addition, 728 control twin pairs were also tested.

TABLE 1

Expected and Observed Twin and Cosib Regressions for Reading Performance

$B_{MZ} = h^2 + c^2_{(t)} = .839$
$B_{DZ} = .5\,h^2 + c^2_{(t)} = .599$
$B_S = .5\,h^2 + c^2_{(s)} = .378$

Note: B_{MZ} = regression of cotwin scores on MZ proband scores; B_{DZ} = regression of cotwin scores on DZ proband scores; B_S = regression of cosib scores on twin or sib proband scores.

Analyses

Table 1 presents the three expected co-relative/proband regressions, namely, cotwins on MZ probands (B_{MZ}), cotwins on DZ probands (B_{DZ}), and cosibs on twin or sib probands (B_S), where $c^2_{(t)}$ is the proportion of observed variance due to environmental influences shared by twin pairs, and $c^2_{(s)}$ is that shared by probands and nontwin siblings. It can be seen that h^2 equals twice the difference between B_{MZ} and B_{DZ}. Further, the difference between B_{DZ} and B_S is a simple function of the difference between shared environmental influences in twin pairs versus those in sibling pairs.

Therefore, to test for differential c^2 between twin pairs and sibling pairs, the following simple extension of the augmented model can be fitted simultaneously to data from probands, their cotwins and cosibs:

$$C = B_3P + B_4R + B_5PR + B_6S + B_7PS + A \quad (4)$$

where PS is the product of the proband's score and the dummy code (+.5 for twin pairs and –.5 for sib or twin–

sib pairs). When Equation 4 is fitted to the data and the five partial regression coefficients are estimated simultaneously, $B_7 = B_{DZ} - B_S$ and, thereby, estimates the extent to which individual differences that are due to shared environmental influences differ in twin pairs versus sib pairs (i.e., $c^2_{(t)} - c^2_{(s)}$). Likewise, $B_5 = 2(B_{MZ} - B_{DZ})$, again estimating h^2, but only from the twin data.

Because truncate selection was employed (DeFries & Gillis, 1991), pairs in which both members met criteria for reading difficulties were double-entered for regression analyses. Standard error estimates and significance tests were adjusted accordingly. All regression models were fitted using SPSS for Unix 17.0.

Results and Discussion

The observed regression coefficients for the discriminant function scores (a composite measure of reading performance) estimated from data of MZ twin pairs, DZ pairs and twin–sib pairs are also presented in Table 1. Doubling the difference between B_{MZ} and B_{DZ} indicates that h^2 = .480. Also, $B_{MZ} - h^2 = c^2_{(t)}$ = .359, and $B_{DZ} - B_S = c^2_{(t)} - c^2_{(s)}$ = .221.

Results of fitting Equation 3 to the twin data only and Equation 4 to the twin–sib data are presented in Table 2. When Equation 3 was fitted to data from twins only, the B_3 and B_5 estimates are $c^2_{(t)}$ = .359 and h^2 = .480, as expected. Similarly, when Equation 4 was fitted to the data from both twin pairs and twin–sib pairs, B_5 was again .480 and B_7 = .221. Although B_7 is only marginally significant (p = .056), this result suggests that shared environmental influences for members of twin pairs (.359) are substantially greater than those for less contemporaneous twin–sib pairs (359 – .221 = .138).

As illustrated, the results obtained by fitting the augmented twin-cosib model to the data may also be obtained by comparing the MZ, DZ, and cosib-twin regressions presented in Table 1. However, fitting Equation 4 to the data obviously provides a more comprehensive analysis of combined twin and twin/cosib data, as well as direct estimates and tests of significance for h^2 and $c^2_{(t)} - c^2_{(s)}$. The multiple regression analysis of twin and sibling data is also more versatile. For example, Equations 1–4 may be easily extended to include various covariates and tests of significance for differential etiology as a function of either dichotomous or continuous variables (Wadsworth, Olson, & DeFries, 2010).

TABLE 2

Results of Fitting the Augmented and Extended-Augmented Models to Twin Data Only and to Data from Both Selected Twin Pairs and Nontwin Siblings

Twins only	
	$B_3 = c^2_{(t)}$ = .359 ± .148, t = 2.42, p = .016
	$B_5 = h^2$ = .480 ± .197, t = 2.43, p = .015
Twins and nontwin cosibs	
	$B_5 = h^2$ = .480 ±.219, t = 2.19, p = .028
	$B_7 = c^2_{(t)} - c^2_{(s)}$ = .221 ±.116, t = 1.91, p = .056

Although fitting Equation 4 to the twin and cosib data facilitates a more comprehensive analysis of the combined data set, it should be noted that, in Table 2, the significance of the h^2 estimate is slightly less (p = .028) than that when Equation 2 is fitted to the twin-only data (p = .015). As recently discussed by Astrom et al. (2011b), the inclusion of data from siblings in twin studies may result in a reduction in power, for example, when shared environmental influences in twin pairs are substantially greater than those in twin–sib pairs. In such cases, Equation 2 should be fitted to the twin data to estimate h^2 and $c^2_{(t)}$, and then Equation 2 should be fitted to the combined twin–sib data to test for the difference between shared environmental influences in twin pairs versus twin–sib pairs. However, when shared environmental influences in twin pairs are not substantially greater than those for sib pairs, power may be increased by the inclusion of data from siblings in twin studies, especially for those with relatively small sample sizes (cf. Astrom et al., 2011).

It should also be noted that Zieleniewski, Fulker, DeFries, & LaBuda (1987) previously extended Equation 2 to analyze data from both twin pairs and sibling pairs. However, their model included two dummy variables (E_T and E_S) to index shared environmental influences for twins and sibling pairs, respectively, rather than the single dummy variable (S) used in Equation 4. Moreover, fitting their extended model to combined twin and sibling data yields separate estimates of $c^2_{(t)}$ and $c^2_{(s)}$, but no direct test of significance for their difference. Although Zieleniewski et al. (1987) fitted subsequent models that equated E_T and E_S to obtain combined estimates of c^2, and then used changes in R^2 for tests of significance, fitting the more parsimonious Equation 4 to combined twin and sib-pair data facilitates direct estimates of $c^2_{(t)} - c^2_{(s)}$ and corresponding significance tests.

Acknowledgments

The Colorado Learning Disabilities Research Center is supported by The Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) grant HD027802. The continued cooperation of the many families participating in the CLDRC, as well as the work of the staff members of these projects is gratefully acknowledged.

References

Astrom, R. L., Wadsworth, S. J., Olson, R. K., Willcutt, E. G., & DeFries, J. C. (2011). DeFries-Fulker analysis of longitudinal reading performance data from twin pairs ascertained for reading difficulties and from their nontwin siblings. *Behavior Genetics, 41*, 660–667.

Astrom, R. L., Wadsworth, S. J., Olson, R. K., Willcutt, E. G., & DeFries, J. C. (2012). Genetic and environmental etiologies of reading difficulties: DeFries-Fulker analysis of reading performance data from twin pairs and their non-twin siblings. *Learning and Individual Differences* (in press).

DeFries, J. C. (1985). Colorado Reading Project. In D. B. Gray & J. F. Kavanagh (Eds.), *Biobehavioral measures of dyslexia* (pp. 107–122). Parkton, MD: York Press.

DeFries, J. C., Filipek, P. A., Fulker, D. W., Olson, R. K., Pennington, B. F., Smith, S. D., & Wise, B. W. (1997). Colorado Learning Disabilities Research Center. *Learning Disabilities: A Multidisciplinary Journal, 8*, 7–19.

DeFries, J. C., & Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavior Genetics, 15*, 467–473.

DeFries, J. C., & Fulker, D. W. (1988). Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellologiae: Twin Research*, *37*, 205–216.

DeFries, J. C., & Gillis, J. J. (1991). Etiology of reading deficits in learning disabilities: Quantitative genetic analysis. In J. E. Obrzut & G. W. Hynd (Eds.), *Neuropsychological foundations of learning disabilities: A handbook of issues, methods and practice* (pp. 29–47). Orlando, Fl: Academic Press.

DeFries, J. C., Olson, R. K., Pennington, B. F., & Smith, S. D. (1991). Colorado Reading Project: Past, present, and future. *Learning Disabilities: A Multidisciplinary Journal, 2*, 37–46.

Dunn, L. M., & Markwardt, F. C. (1970). *Examiner's manual: Peabody Individual Achievement Test.* Circle Pines, MN: American Guidance Service.

Nichols, R.C., & Bilbro Jr., W.C. (1966). The diagnosis of twin zygosity. *Acta Geneticae Medicae et Gemellologiae, 16,* 265–275.

Wadsworth, S. J., Olson, R. K., & DeFries, J. C. (2010). Differential genetic etiology of reading difficulties as a function of IQ: An update. *Behavior Genetics, 40,* 751–758.

Zieleniewski, A. M., Fulker, D. W., DeFries, J. C., & LaBuda, M. C. (1987). Multiple regression analysis of twin and sibling data. *Personality and Individual Differences, 8,* 787–791.