

# Genetic Etiologies of Comorbidity and Stability for Reading Difficulties and ADHD: A Replication Study

Sally J. Wadsworth,<sup>1</sup> John C. DeFries,<sup>1</sup> Erik G. Willcutt,<sup>1,2</sup> Bruce F. Pennington,<sup>1,3</sup> and Richard K. Olson<sup>1,2</sup>

<sup>1</sup>Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, USA

<sup>2</sup>Department of Psychology and Neuroscience, University of Colorado, Boulder, Colorado, USA

<sup>3</sup>Department of Psychology, University of Denver, Denver, Colorado, USA

Because of recent concerns about the replication of published results in the behavioral and biomedical sciences (Ioannidis, *PLoS Medicine*, Vol. 2, 2005, p. e124; Open Science Collaboration, *Science*, Vol. 349, 2015, p. 943; Pashler & Wagenmakers, *Perspectives on Psychological Science*, Vol. 7, 2012, pp. 528–530), we have conducted a replication of our recently published analyses of longitudinal reading performance and attention deficit-hyperactivity disorder data from twin pairs selected for reading difficulties (Wadsworth et al., *Twin Research and Human Genetics*, Vol. 18, 2015, pp. 755–761). Results obtained from univariate and bivariate (DeFries & Fulker, *Behavior Genetics*, Vol. 15, 1985, pp. 467–473; *Acta Geneticae Medicae et Gemellologiae: Twin Research*, Vol. 37, 1988, pp. 205–216) analyses of data from a subset of twin pairs tested in the International Longitudinal Twin Study of Early Reading Development at post-4th grade, and its continuation into high school at post-9th grade, were compared to those from our previous report. Similar measures of reading performance, the same measures of inattention and hyperactivity/impulsivity, and similar selection criteria were used in the two studies. In general, the patterns of results obtained from these two independent studies were highly similar. Thus, these results clearly illustrate the principle that findings from studies in quantitative behavioral genetics often replicate (Plomin et al., *Perspectives on Psychological Science*, Vol. 11, 2016, pp. 3–23).

■ **Keywords:** reading difficulties, ADHD, longitudinal, twin study, comorbidity, genetic

We recently reported results of the first longitudinal twin study of reading difficulties (RD) and attention deficit-hyperactivity disorder (ADHD) symptom dimensions in the Colorado Learning Disabilities Research Center (CLDRC; DeFries et al., 1997), from a sample of twin pairs selected for RD (Wadsworth et al., 2015). The purpose of that study was to assess the etiology of the stability of RD as well as the etiology of comorbidity between RD and ADHD symptom dimensions both contemporaneously and longitudinally, using univariate and bivariate DeFries–Fulker (DF) analyses (DeFries & Fulker, 1985; 1988). Reading composite data based on the Reading Recognition, Reading Comprehension, and Spelling subtests of the Peabody Individual Achievement Test (PIAT; Dunn & Markwardt, 1970) and ADHD symptom dimensions (inattention [IN] and hyperactivity/impulsivity [H/I]) from the Disruptive Behavior Rating Scale (DBRS; Barkley & Murphy, 1998) were analyzed from twin pairs in which at

least one member met proband criteria for RD at initial assessment, and in which both members of the pair had data from a follow-up assessment approximately 5 years later. In order for an individual to be classified as RD, he or she was required to have a positive history for reading problems and be classified as affected by scores on the reading composite. Additional diagnostic criteria included a verbal or performance IQ score of at least 85 on the Wechsler Intelligence Scale for Children — Revised (WISC-R; Wechsler, 1974) or the Wechsler Adult Intelligence Scale — Revised (WAIS-R; Wechsler, 1981); no evidence of neurological problems;

---

RECEIVED 14 June 2016; ACCEPTED 13 September 2016. First published online 17 October 2016.

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](mailto:sally.wadsworth@colorado.edu)

and no uncorrected visual or auditory acuity deficits. The subjects ranged in age from 7.7 to 20.5 years (average age of 11.6 years) at initial assessment, and from 12.6 to 26.6 years (average age of 16.2 years) at follow-up.

The genetic etiologies of RD and of the comorbidity between RD and ADHD at the initial measurement occasion were assessed by DF analyses of data from 767 twin pairs for the univariate analysis of RD and 345 pairs for the bivariate analyses of RD and ADHD. In addition, data were analyzed from 94 twin pairs in which at least one member of each pair met proband criteria for RD and for whom reading data were available at both measurement occasions, as well as from 88 twin pairs that also had ADHD data at follow-up. Results of these analyses indicated that more than 60% of the proband deficit in reading at initial assessment was due to genetic influences, and that reading deficits at follow-up were substantially due to these same genetic influences ( $\text{Biv } h_g^2 = 0.79 \pm 0.22$ ). Results of bivariate DF analyses of initial reading and both initial and follow-up symptoms of IN indicated that genetic influences accounted for 60% of the contemporaneous relationship and approximately two-thirds of the longitudinal relationship ( $\text{Biv } h_g^2 = -0.68 \pm 0.33$ ). In contrast, bivariate  $h_g^2$  estimates for the comorbidity between initial reading and both contemporaneous and follow-up H/I symptoms were small and non-significant (Wadsworth et al., 2015). In summary, our previous findings based on analyses of data from the CLDRC selected sample indicated strong genetic influences on RD at initial assessment, as well as on comorbidity, between RD and IN at initial and follow-up assessments (see Table 1).

Recently, it has been noted that many statistically significant findings in the behavioral sciences have not replicated (Pashler & Wagenmakers, 2012; Plomin et al., 2016). A recent attempt to replicate findings of 100 such studies found that 64% failed to replicate (Open Science Collaboration, 2015). In an attempt to replicate 17 brain-behavior studies, Boekel et al. (2015) found that none replicated. Results of attempts to replicate medical findings have been similarly discouraging, with five of six non-randomized designs failing to replicate (Ioannidis, 2005). These and similar results have led to claims that 85% of research resources are wasted (Macleod et al., 2014).

The International Longitudinal Twin Study of Early Reading Development (ILTS; Byrne et al., 2006) and its continuation into high school provide an exceptional opportunity to conduct a replication of our previous study using similar measures and selection criteria, and exactly the same analyses. To accomplish this, we chose those measurement occasions (post-4th grade and post-9th grade) that corresponded most closely in age to the mean ages at initial and follow-up assessments in the CLDRC (11.6 and 16.2 years, respectively) and selected those twin pairs in which at least one member of the pair had RD at post-4th-grade assessment (average age 10.5 years) and follow-up data at post-9th grade (average age 15.5 years).

## Methods

### Participants

Subjects in the current study are participants in the ongoing ILTS (Byrne et al., 2006) that includes twins from Australia, the United States, and Scandinavia. However, the subset of twins whose data were used in the current study include only those participating in the U.S. (Colorado) study. Twins were recruited from birth records, and zygosity was determined from DNA extracted from cheek swabs, or in a minority of cases (28%, most of whom were clearly fraternal) from selected items from the Nichols and Bilbro (1966) questionnaire. All twins were learning to read English at entrance into the study. Those twin pairs in which at least one member of the pair had a composite reading score at least one standard deviation below the full sample mean at post-4th grade, and scores on either the WPPSI Vocabulary or Block Design at entrance of no more than one standard deviation below the sample mean at entry into the study were selected for analyses. The subsample selected for RD at the end of 4th grade consisted of 86 twin pairs, 38 monozygotic (MZ; i.e., identical), and 48 same-sex dizygotic (DZ; i.e., fraternal). By post-9th grade, the sample consisted of 34 MZ and 46 DZ pairs.

### Procedure and Measures

The measures included in the present analyses are from larger test batteries that were administered in the ILTS in the summer after each school year. Testing at each time point was conducted in a single session in the twins' homes or schools. Two testers separately assessed each twin at the same time. The following measures were included in the current analyses.

**Reading.** The Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999), Sight Word Efficiency, as well as the Woodcock-Johnson Word ID and Passage Comprehension (Woodcock et al., 2001) were administered at both post-4th grade and post-9th grade.

**ADHD.** IN and H/I were measured using nine items relating to IN and nine relating to H/I from the parent and teacher versions of the DBRS (Barkley & Murphy, 1998). These items have been shown to be a valid and reliable measure of ADHD symptoms in children (Lahey et al., 2004; Willcutt et al., 2007).

**Verbal and performance IQ.** WPPSI Vocabulary, assessed at entry into the study at pre-Kindergarten, was used as a proxy for verbal IQ, and Block Design was used as a proxy for performance IQ.

### Analyses

**Multiple regression analysis of twin data.** Although qualitative analysis such as a comparison of concordance rates is appropriate as a test for genetic etiology of a

**TABLE 1**Results of Univariate and Bivariate DF Analyses of Two Studies<sup>a</sup>

Analysis	CLDRC results [CI]	ILTS results [CI]
Univariate DF — Initial reading	$h_g^2 = 0.63 \pm 0.06, p < 1.1 \times 10^{-16}$ [0.50, 0.75]	$h_g^2 = 0.69 \pm 0.23, p \leq .0015$ [0.23, 1.14]
Bivariate DF — Initial and follow-up reading	Biv $h_g^2 = 0.79 \pm 0.22, p < .0003$ [0.35, 1.23]	Biv $h_g^2 = 0.72 \pm 0.31, p \leq .013$ [0.11, 1.34]
Bivariate DF — Initial reading and initial IN	Biv $h_g^2 = -0.60 \pm 0.15, p < 6.9 \times 10^{-5}$ [-0.90, -0.20]	Biv $h_g^2 = -0.40 \pm 0.30, p \geq .09$ [-1.00, 0.20]
Bivariate DF — Initial reading and follow-up IN	Biv $h_g^2 = -0.68 \pm 0.33, p < .02$ [-1.33, -0.04]	Biv $h_g^2 = -0.33 \pm 0.31, p \geq .14$ [-0.95, 0.28]
Bivariate DF — Initial reading and initial H/I	Biv $h_g^2 = -0.020 \pm 0.15, p > .097$ [-0.50, 0.10]	Biv $h_g^2 = -0.29 \pm 0.32, p \geq .18$ [-0.92, 0.35]
Bivariate DF — Initial reading and follow-up H/I	Biv $h_g^2 = 0.11 \pm 0.37, p > .38$ [-0.62, 0.84]	Biv $h_g^2 = -0.27 \pm 0.34, p \geq .21$ [-0.95, 0.40]

Note: <sup>a</sup>All  $p$  values are one-tailed. DF = DeFries–Fulker.

dichotomous variable, such as diagnosis of an illness or behavioral disorder, RD and ADHD symptoms occur on a continuum, with somewhat arbitrary cut-off points designating an individual as ‘affected’ or ‘unaffected’. Therefore, DeFries and Fulker (1985) proposed a multiple regression analysis of twin data to assess the etiology of extreme scores on a continuous measure. A basic model was proposed in which a co-twin’s score is predicted from the proband’s score on the selected trait and the coefficient of relationship (1.0 and 0.5 for identical and fraternal twin pairs, respectively) such that

$$C = B_1 P + B_2 R + A, \quad (1)$$

where  $C$  symbolizes the co-twin’s score,  $P$  is the proband’s score,  $R$  is the coefficient of relationship, and  $A$  is the regression constant.  $B_1$  is the partial regression of the co-twin’s score on the proband’s score, a measure of average MZ and DZ twin resemblance,  $B_2$  is the partial regression of the co-twin’s score on the coefficient of relationship and equals twice the difference between the MZ and DZ co-twin means after covariance adjustment for any difference between MZ and DZ proband means. As a result,  $B_2$  provides a direct test for genetic etiology. Further, when the data are appropriately transformed prior to multiple-regression analysis (i.e., each score is expressed as a deviation from the mean of the unselected population and then divided by the difference between the proband and population means),  $B_2 = h^2 g$ , an index of the extent to which the average deficit of the probands is due to genetic influences (DeFries & Fulker, 1988). For the current analyses, the unselected population is represented by the full population sample of twin pairs at each assessment.

**Etiologies of stability and comorbidity.** The DF multiple regression model may be extended to assess the relationship between two different phenotypes or the same phenotype at two different time points. For example, to assess the etiology of stability between deficits in reading performance at the two time points, the following bivariate extension of the basic regression model was fitted to proband reading scores at initial assessment and co-twins’ scores at follow-up:

$$C_y = B_1 P_x + B_2 R + A, \quad (2)$$

where  $C_y$  is the co-twin’s score at follow-up ( $Y$ ) and  $P_x$  is the proband’s score at initial assessment. In the bivariate case,  $B_1$  is the partial regression of the co-twin’s reading score at follow-up ( $Y$ ) on the proband’s initial reading score ( $X$ ), a measure of the average MZ–DZ cross-variable twin resemblance, or the extent to which co-twin scores on  $Y$  are related to proband scores on  $X$  (in this case, reading) across zygosity.  $B_2$  is the partial regression of the co-twin’s  $Y$  score on the coefficient of relationship. When the data are appropriately transformed,  $B_2 = h_x h_y rG(xy)$ , an index of the extent to which the proband deficit on  $X$  is due to genetic factors that also influence scores on  $Y$ , that is, ‘bivariate heritability’ (Light & DeFries, 1995).  $rG(xy)$  is the genetic correlation, an index of the degree to which individual differences in two variables are due to the same genetic influences. Thus, Equation (2) can also be applied to assess the genetic etiologies of both contemporaneous and longitudinal comorbidities between RD and ADHD symptom dimensions.

In the current study, the etiology of reading deficits at 4th grade was assessed, as well as their longitudinal stability between 4th grade and 9th grade. In addition, both the contemporaneous relations between 4th-grade reading and 4th-grade IN and H/I, and the longitudinal relations between 4th-grade reading and 9th-grade IN and H/I were assessed. In order to provide strictly parallel analyses to the CLDRC analyses, subjects were not reselected at 9th grade.

## Results

Table 1 presents results of both the previously published analyses of data from the CLDRC and those of the current study. Although there are some relatively minor differences between the results, the overall pattern of results and indeed most estimates are highly similar. In both studies, the heritability of the group deficit in reading at initial assessment is greater than 60%. Also, in both studies, genetic influences on stability of the reading deficit are greater than 70%. Although the bivariate heritability for initial reading and IN in the CLDRC (-0.60) is larger than the corresponding estimate for the ILTS (-0.40), and the difference is even greater for the bivariate heritability of initial reading and follow-up IN (-0.68 vs. -0.33), their confidence intervals overlap

substantially. In addition, bivariate heritabilities for initial reading and both initial and follow-up H/I are somewhat lower than the corresponding estimates for IN in both studies.

## Discussion

The failure of many findings in the behavioral and biomedical sciences to replicate may have many possible causes, including differences in populations, ages of subjects, measures, diagnostic criteria, and so forth. Thus, the current study, based on analyses of data from a selected subset of a population sample, has attempted to replicate our previous findings from a selected sample using identical analyses, as well as highly similar measures and diagnostic criteria. Sample sizes differed depending on the measures analyzed and samples from which subjects were drawn, but were similar for the bivariate analyses in the two studies. Results obtained from DF analyses indicated that reading deficits at initial assessment and their stability are due substantially to genetic influences in both studies. Also, results of both studies suggested that genetic influences on the comorbidity between initial reading and IN were greater than those on the comorbidity between initial reading and H/I, both contemporaneously and longitudinally.

As indicated by their relatively large confidence intervals, the differences between the CLDRC and ILTS bivariate heritabilities for initial reading and IN may only be due to chance. However, these differences could also be due in part to some minor differences in sample and procedure. First, the CLDRC sample is a selected sample. Although a subset of subjects was selected for these analyses in the ILTS, the selection criteria were not exactly the same, and indeed could not be the same due to differences in measures administered. Second, although the mean ages of subjects at each measurement occasion were similar, there was a wide range of ages at both measurement occasions in the CLDRC, with the range of ages at initial assessment from 7.7 to 20.5 years of age, and at follow-up from 12.6 to 26.6 years of age, whereas in the ILTS, all subjects were post-4th grade and post-9th grade, with little range in age at each assessment. Further, the measures of reading also differed somewhat for the two samples.

## Conclusions

Our previous findings of substantial genetic influences for reading deficits and their longitudinal stability are clearly replicated in this independent analysis of twin data. Although the bivariate  $h_g^2$  estimates between RD and IN are somewhat lower in this replication study, the bivariate heritability estimates between reading deficits and H/I are relatively low in both studies. Nevertheless, the minor differences between these results clearly illustrate the need for standardization of procedures and measures, as well as for the importance of replication.

## Acknowledgments

The continued cooperation of the many families and schools participating in the CLCRC and ILTS, as well as the work of the staff members of these projects, is gratefully acknowledged. The Colorado Learning Disabilities Research Center is supported by grant HD027802; the U.S. Sample of the International Longitudinal Twin Study of Early Reading Development is supported by grant HD038526, and its continuation into high school, Etiology and Neuropsychology of Math, Reading, ADHD, and Their Covariation by grant HD068728, all from the Eunice Kennedy Shriver Center of the National Institute of Child Health and Human Development (NICHD).

## Conflict of Interest

None.

## Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

## References

- Barkley, R. A., & Murphy, K. (1998). *Attention-deficit hyperactivity disorder: A clinical workbook* (2nd ed.). New York: Guilford Press.
- Boekel, W., Wagenmakers, E. J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, 66, 115–133.
- Byrne, B., Olson, R. K., Samuelsson, S., Wadsworth, S., Corley, R., DeFries, J. C., & Willcutt, E. (2006). Genetic and environmental influences on early literacy. *Journal of Research in Reading*, 29, 33–49.
- 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.
- Dunn, L. M., & Markwardt, F. C. (1970). *Peabody individual achievement test*. Circle Pines, MN: American Guidance Service.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124.
- Lahey, B. B., Pelham, W. E., Loney, J., Kipp, H., Ehrhardt, A., Lee, S. S., ... Massetti, G. (2004). Three-year predictive validity of DSM-IV attention deficit hyperactivity disorder in

- children diagnosed at 4-6 years of age. *American Journal of Psychiatry*, 161, 2014–2020.
- Light, J. G., & DeFries, J. C. (1995). Comorbidity of reading and mathematics disabilities: Genetic and environmental etiologies. *Journal of Learning Disabilities*, 28, 96–106.
- Macleod, M., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J., ... Glasziou, P. (2014). Biomedical research: Increasing value, reducing waste. *Lancet*, 383, 101–104.
- Nichols, R. C., & Bilbro, W. C. (1966). The diagnosis of twin zygosity. *Acta Genetica et Statistica Medica*, 16, 265–275.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, 943.
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528–530.
- Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science*, 11, 3–23.
- Torgesen, J., Wagner, R., & Rashotte, C. A. (1999). *A test of word reading efficiency (TOWRE)*. Austin, TX: PRO-ED.
- Wadsworth, S. J., DeFries, J. C., Willcutt, E. G., Pennington, B. F., & Olson, R. K. (2015). The Colorado Longitudinal Twin Study of Reading Difficulties and ADHD: Etiologies of comorbidity and stability. *Twin Research and Human Genetics*, 18, 755–761.
- Wechsler, D. (1974). *Examiner's manual: Wechsler intelligence scale for children - revised*. New York: The Psychological Corporation.
- Wechsler, D. (1981). *Examiner's manual: Wechsler adult intelligence scale - revised*. New York: The Psychological Corporation.
- Willcutt, E. G., Betjemann, R. S., Pennington, B. F., Olson, R. K., DeFries, J. C., & Wadsworth, S. J. (2007). Longitudinal study of reading disability and attention-deficit/hyperactivity disorder: Implications for education. *Mind, Brain, and Education*, 4, 181–192.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside Publishers.