

Modeling Rater Disagreement for ADHD: Are Parents or Teachers Biased?

Christie A. Hartman · Soo H. Rhee · Erik G. Willcutt ·
Bruce F. Pennington

Published online: 27 February 2007
© Springer Science+Business Media, LLC 2007

Abstract The present study is the first to utilize twin modeling to examine whether parent-teacher disagreement for ADHD ratings is due to parent or teacher bias, or due to raters observing different but valid ADHD behaviors. A joint analysis was conducted with 106 twin pairs, including twins selected for ADHD and control twin pairs. Total ADHD scores were analyzed using multiple rater models that estimate genetic and environmental contributions common to both raters and unique to each rater. Results suggest that 1) disagreement in ADHD ratings is strongly due to parents and teachers observing different ADHD behaviors, some of which is valid and some of which is due to bias, and 2) parents may be more biased than teachers in their ADHD ratings.

Keywords ADHD · Rater bias · Twin modeling · Rater disagreement

Introduction

Attention-deficit/hyperactivity disorder (ADHD), characterized by persistent and maladaptive levels of inattention

and/or hyperactivity-impulsivity, occurs in the general population at a rate of approximately 5% (APA., 1994). In order to assess ADHD symptoms, researchers rely upon behavioral ratings from those who interact with the child on a regular basis, often parents or teachers. However, parents and teachers only partially agree on their ratings of a particular child's ADHD symptoms – correlations between mother and teacher ratings range from .37 to .49 (DuPaul, Power, Anastopoulos, & Reid, 1998; Gadow & Sprafkin, 1999; Sprafkin, Volpe, Gadow, Nolan, & Kelly, 2002; Willcutt, Hartung, Lahey, Pelham, & Loney, 1999). Do these rater differences occur because parents or teachers are somehow biased in their ratings, or because they are observing different behaviors?

Parent-teacher disagreement with ADHD ratings can have ramifications for genetically informed studies such as twin, linkage, and association studies. The results of these studies partially depend on how one defines the ADHD phenotype. Thus, if parents and teachers differ in their ADHD ratings, the ADHD phenotype may differ depending on who rates the child, thereby influencing results. An example of this has been observed in twin studies, where parent-teacher rating differences have been well-documented. Twin studies evaluate phenotypic similarity between monozygotic (MZ) and dizygotic (DZ) twins – greater similarity in MZ twins compared to DZ twins indicates genetic etiology. A common finding with ADHD (but not other externalizing disorders) is that DZ twin correlations from parent ratings are often lower than those from teacher ratings; sometimes the DZ correlations from parent ratings are close to zero, or even negative. This pattern of twin correlations has been observed in different samples (Eaves, Silberg, Meyer, & Maes, 1997; Sherman, Iacono, & McGue, 1997; Thapar, Hervas, & McGuffin, 1995), with DSM- and non-DSM-based ADHD measures (Eaves, Silberg, Meyer, & Maes, 1997; Silberg et al., 1996),

C. A. Hartman (✉)
Department of Psychiatry, University of Colorado at Denver and
Health Sciences Center, 4200 East 9th Ave.,
Box C-268-35, Denver, CO 80262
e-mail: christie.hartman@colorado.edu

C. A. Hartman · S. H. Rhee · E. G. Willcutt
Institute for Behavioral Genetics, University of Colorado
at Boulder

B. F. Pennington
Department of Psychology, University of Denver

and with hyperactivity only (Goodman & Stevenson, 1989; Silberg et al., 1996; Thapar, Hervas, & McGuffin, 1995).

DZ twin correlations that are less than half the MZ twin correlations can indicate non-additive genetic (dominance) effects. However, the fact that this pattern is only seen with parent ratings and not teachers suggests a rater effect. Low DZ correlations relative to MZ correlations can also indicate sibling contrast effects, and twin models incorporating these effects provide better fit for parent-rated ADHD (Eaves, Silberg, Meyer, & Maes, 1997; Silberg et al., 1996; Thapar, Hervas, & McGuffin, 1995). These contrast effects may be due to rater bias (Eaves, Silberg, Meyer, & Maes, 1997; Simonoff et al., 1998), where parents tend to exaggerate differences between their DZ twins because they are contrasting the twins with one another, as opposed to having many children with whom to compare the twins, as teachers do (Eaves, Silberg, Meyer, & Maes, 1997). Thus, it is possible that rater disagreement is somewhat due to bias in parent ratings.

Although rater bias is a plausible explanation for some of the differences in how parents and teachers rate ADHD, another possibility is that parents and teachers rate twins differently because they are rating different behaviors. A DSM-IV diagnosis of ADHD requires that a child exhibit significant impairment in at least two domains (typically home and school), as the child may exhibit different symptoms at home and school because each setting presents different demands. Thus, when parents and teachers rate twins on their ADHD symptoms, are they assessing the same phenotype? Thapar et al. (2000) examined parent and teacher ADHD ratings in a twin analysis, and found that both parent- and teacher-rated ADHD shared a common genetic origin but that teacher-rated ADHD had its own unique genetic and environmental contributions. Their results suggest that parents and teachers may be assessing somewhat different phenotypes.

Through twin modeling, we can go beyond the limitations of examining twin correlations and answer this question more directly using multiple rater models, which allow us to estimate the genetic and environmental contributions common to both raters as well as the genetic and environmental contributions unique to each rater (Hewitt, Silberg, Neale, Eaves, & Erickson, 1992; Neale & Cardon, 1992). Using these models, we can estimate the presence and extent of rater bias effects in parents and teachers, and whether parents and teachers are examining different (but valid) ADHD phenotypes. The simpler of these models, called the Rater Bias model, assumes both raters are assessing a single phenotype. The second model, known as the Psychometric model, is similar to the Rater Bias model except that it also estimates genetic contributions to a phenotype that are unique to each rater, which, if significant, would suggest that the raters are assessing different phenotypes to some extent. Van der Valk, van den Oord, and Boomsma (2001) utilized these multiple rater models to examine mother and father ratings of CBCL

internalizing and externalizing in Dutch twins. They found that the Psychometric model provided better fit to the data, and although mothers and fathers were largely assessing the same phenotype, there was evidence for a component that was unique to each rater, but valid. They also detected modest evidence for rater bias.

In order to better understand parent-teacher disagreement with ADHD ratings, we analyzed parent- and teacher-rated ADHD using the Rater Bias and Psychometric models. To our knowledge, our study is the first to utilize these models with parent and teacher ratings of ADHD symptoms. We sought to examine whether parent or teacher ratings show evidence of rater bias, and whether parents and teachers are assessing the same ADHD phenotype.

Method

Sample

Recruitment

Participants for this study are part of the Colorado Learning Disabilities Research Center (CLDRC) twin study, an ongoing study of the etiology of learning disabilities, ADHD, and other related disorders (DeFries et al., 1997). In collaboration with 27 local school districts, parents of all twins between the ages of 8 and 18 were contacted by letter and invited to participate in the study. Approximately 35% of the families who were contacted agreed to participate in the initial screening procedure. After obtaining parental consent, parents and teachers were asked to complete the *Disruptive Behavior Rating Scale* (Barkley & Murphy, 1998) to assess symptoms of DSM-IV ADHD (APA., 1994). If one of the twins met symptom criteria for any DSM-IV ADHD subtype based on parent or teacher ratings, the twin pair was recruited for the twin study. In addition, twin pairs in which one twin had a history of reading difficulties were recruited independent of the procedure to ascertain the sample of twins with ADHD, and a third group of twins without ADHD or reading difficulties was recruited from the same school districts as a comparison sample. These procedures yielded a community-based sample that is enriched for ADHD and learning difficulties. Approximately 95% of the families in the screening sample agreed to participate in the larger study if invited. Participants with documented brain injury or IQ scores below 75 were excluded from the study.

The representativeness of the sample that agreed to participate was indirectly examined by comparing the characteristics of the families who participated in the study to the demographic information for each school that is available in the public record. These comparisons indicated that the families who agreed to participate in the screening were drawn

proportionally from the schools in each district, and were representative of the overall population of each district in terms of gender ratio and ethnicity.

Sample for the present analyses

The total sample included 119 MZ twin pairs (38 selected and 81 control) and 190 DZ twin pairs (85 selected and 105 control). Although the overall sample ranged in age from 8 to 18 years, recruitment was weighted toward individuals between 8 and 13 years of age (85% of the overall sample) to facilitate follow-up analyses in a separate longitudinal component of the study. Therefore, the mean age of the present sample was 10.6 years ($SD = 2.5$). The ethnicity of the participants included in these analyses (82% Caucasian, 9% Hispanic, 4% African American, 5% other ethnicity) is consistent with the ethnic breakdown of the overall CLDRC sample and the total population of students in the school districts from which the twins were recruited. The overall sample was comprised of 52% females and 48% males – however, there were more males in the selected group (62%) than in the control group (41%), as would be expected.

Measures

Symptoms of ADHD

The Disruptive Behavior Rating Scale (DBRS; Barkley & Murphy, 1998) was used to obtain parent and teacher ratings of the 18 symptoms of DSM-IV ADHD. Because maternal ratings were available for more participants than were paternal ratings (95% vs. 73% of the sample), and because we wanted to reduce potential error created by multiple informant sources, only maternal ratings were used for the analyses described in this study. Each item on the DBRS is rated on a four-point scale (*never or rarely, sometimes, often, and very often*). Previous results from this sample and others indicate that parent and teacher ratings on the DBRS or other similar scales are internally consistent ($\alpha = .92-.96$) and have adequate to high test-retest reliability ($r = .59-.89$; e.g., DuPaul, Power, Anastopoulos, & Reid, 1998; Willcutt, Chhabildas, & Pennington, 2001).

Analysis

ADHD phenotype

ADHD diagnoses for selected twins were determined by combining parent and teacher reports using the “or-rule,” which codes each ADHD symptom as positive if it is endorsed by either the parent or the teacher (Piacentini, Cohen, & Cohen, 1992). This procedure was utilized in the DSM-IV field trials (Lahey et al., 1994), and is a standard method for

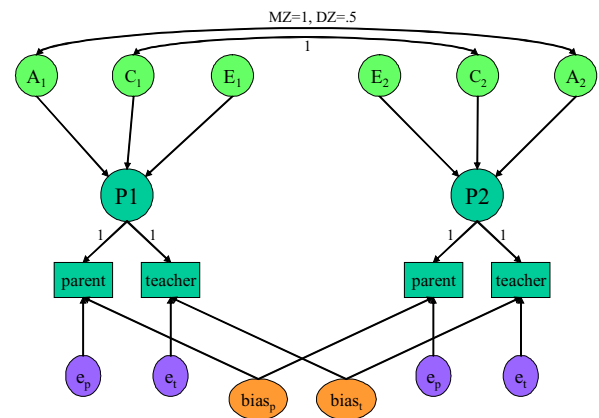


Fig. 1 Rater Bias Model

Note. P1 = twin 1 phenotype, P2 = twin 2 phenotype, p = parent, t = teacher, A = genetic effects, C = shared environmental effects, E = non-shared environmental effects, e = measurement error, $bias$ = rater bias

combining data from different informants (Costello et al., 1988; Simonoff et al., 1997).

For all analyses, scores for all DSM-IV ADHD symptoms were summed for each participant. The total symptom scores were log transformed to approximate normality.

Multiple rater models

In order to examine both parent and teacher ratings simultaneously and analyze rater effects, Rater Bias and Psychometric models (Hewitt et al., 1992; Neale & Cardon, 1992) were fit to the data (Figs 1 and 2). In the basic twin model with one informant source, the variance in scores is decomposed into genetic (A), shared environmental (C), and unique environmental (E) sources. A greater MZ than DZ correlation is evidence of genetic influences. An MZ correlation that is less than 1.0 is evidence of nonshared environmental influences. A DZ twin correlation that is greater than half of the MZ twin correlation is evidence of shared environmental influences. In the multiple rater models utilized in this study, the cross-correlation (the correlation between the twins rated by different raters) is decomposed into genetic, shared environmental, and unique environmental sources in the same way.

In the Rater Bias model (Fig. 1), the ADHD phenotypes for each twin, measured by parent and teacher ratings, are decomposed into three sources of variance common to both raters: additive genetic effects (A), shared environmental effects (C), and non-shared environmental effects (E). These parameters represent the shared rater view. The ratings are also decomposed into sources of variance unique to each rater, including rater bias ($bias$) and measurement error (e); these parameters represent rater disagreement (Bartels et al., 2003).

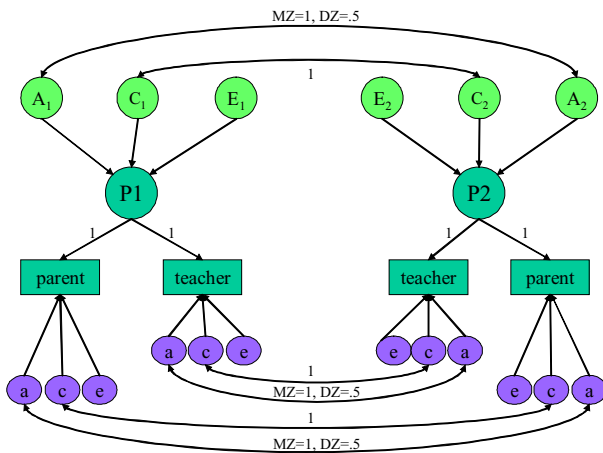


Fig. 2 Psychometric Model
 Note. P1 = twin 1 phenotype, P2 = twin 2 phenotype, p = parent, t = teacher, A = genetic effects, C = shared environment effects, E = non-shared environment effects, a = rater-specific genetic effects, c = rater-specific shared environment (rater bias), e = rater-specific non-shared environment (measurement error)

In the Psychometric model (Fig. 2), the phenotypic variance for each twin is also decomposed into A, C, and E common to both raters, as well as genetic, shared environment, and non-shared environment factors unique to each rater (a, c, and e, respectively). Shared environmental factors unique to a rater (c) represents an estimate of rater bias, and (c) is equivalent to “bias” in the Rater Bias model. Rater bias (c) represents systematic response styles seen among parents or among teachers, and is suspected when (c) estimates are higher than expected. Finally, non-shared environmental factors unique to a rater (e) can also represent measurement error, labeled (e) in the Rater Bias model as well. A, C, and E (common rater view) are estimated from the twin cross-correlations, and the unique rater view parameter estimates (a, c, and e) are calculated from the difference between the variance shared between raters (the common view) and the total variance.

Once unique rater view parameters are estimated, the parameter for genetic effects unique to a rater (a) is one way that allows us to examine whether rater differences (unique rater view) are valid or biased. If behavior uniquely rated by parents and teachers is influenced by genes, then it is assumed that the rater is assessing valid behaviors in the child because rater bias (c) and measurement error (e) cannot produce the systematic effects needed to produce evidence of unique genetic effects (Bartels et al., 2003). Thus, if significant, (a) suggests that, to some measurable extent, each rater is observing a unique but valid phenotype. In addition, shared environmental effects *common* to both raters (C) are indicated if twin correlations are larger than zero once genetic effects are controlled for; thus, when shared environmental effects *unique* to a rater (c) are greater than zero, this

suggests rater bias, as the rater will bias the ratings of both twins in the same way, and in a way that differs from the other rater.

Rater Bias and Psychometric models are nested, in that they are the same except for the Psychometric model estimating one additional parameter: genetic effects unique to a rater (a). It is important to note that for these two models to be nested, a restricted version of the Rater Bias model must be utilized, which simply constrains both pathways from the latent phenotypic variable to the observed variables to one (please see Hewitt et al., 1992, for more details).

Model fitting

All analyses were conducted using Mx (Neale, 1997). When fitting raw data in Mx, the overall fit of each model is expressed as $-2LL$ (minus twice the log likelihood). When comparing two nested models, the absolute difference in $-2LL$ is distributed as χ^2 and can then be tested for significance using the χ^2 distribution, with degrees of freedom equal to the difference in degrees of freedom between the two models (Kline, 1998). If the less complex model (estimating fewer parameters) yields no significant decrement in fit (as indicated by a non-significant change in $-2LL$), that model then represents the “best fitting” model among that pair of models. Model fit was also evaluated using Akaike’s information criterion (AIC; Akaike, 1987), which takes both model fit and parsimony into account. The formula for AIC is $-2LL$ minus twice the degrees of freedom. When comparing two models, the model with the lowest AIC best represents the data.

Modeling selected and control samples

Our sample is comprised of probands selected for ADHD as well as a separate control group. Selected samples create some analytic challenges; for example, compared to using population samples, selected samples provide greater power due to higher prevalence of a disorder in the sample, but results from selected samples may not generalize to the population. Moreover, selection tends to underestimate twin correlations and variances, resulting in biased estimates for genetic and environmental parameters. In order to account for selection effects in our sample, we performed a joint analysis of data from a selected group as well as an unselected (control) group of twins. This method provides the greater power of a selected sample while allowing for the results to be generalized to the population.

We used an approach based on methods devised by Pearson (1902) and Aitken (1934). Briefly, the differences in prevalence between the selected and control samples provide an estimate of the magnitude of selection. Rather

than lumping selected and control groups together, this method estimates selected proband means and variances separately from those of controls. Once the magnitude of selection was estimated, we adjusted the selected twin variance/covariance expectations for the effects of selection. For greater detail on this method, please see Stallings et al. (1997). It should be noted that while modeling selected and control samples separately will produce less biased twin estimates by estimating the magnitude of selection, the controls used in our study were selected for not having ADHD, and thus don't necessarily represent a true population sample.

Results

Parent-teacher agreement for ADHD

The agreement between parent and teacher ratings of ADHD was moderate at $r = 0.52$. This correlation is slightly higher than those reported in other studies (DuPaul et al., 1998; Gadow & Sprafkin, 1999; Sprafkin et al., 2002; Thapar et al., 2000; Willcutt et al., 1999).

Multiple rater model fitting

Overall model fit

Because some of the twin pairs were rated by more than one teacher, which could result in biased estimates for teachers, we removed those twin pairs from the primary analyses. This resulted in a sample of 106 twin pairs (36 selected and 70 controls) with similar gender and ethnic breakdown as found with the larger sample. The Psychometric Model provided better fit to the data than the Rater Bias Model, as shown by the significance of the χ^2 difference test (Table 1). Because the parameter for shared environmental factors common to both raters (C) was estimated at zero, it was dropped from both models. The Psychometric model also showed improved fit after accounting for its greater complexity, as indicated by its lower AIC value. These results suggest that parents and teachers are observing somewhat different ADHD phenotypes.

Table 1 Model fitting results

	- 2LL (df) ^a	AIC ^b	χ^2 diff. (df) ^c	p (χ^2 diff.)
Rater Bias AE	286.4 (373)	- 459.6		
Psychometric AE	275.8 (371)	- 466.2	6.60 (2)	.037

^aminus twice the log likelihood (degrees of freedom).

^bAkaike's Information Criterion.

^c χ^2 difference test is between the two models.

Parameter estimates

When examining the standardized parameter estimates for the better-fitting Psychometric model (Table 2), the proportion of the total variance in ADHD scores due to a common rater view was 37% for parents and 30% for teachers. Likewise, the proportion of the total variance in scores due to unique rater view was larger at 63% and 70% for parents and teachers, respectively. This suggests that while parents and teachers are observing somewhat similar ADHD behaviors in these children, to a larger extent they are observing different behaviors.

Further examination of the parameter estimates showed that the unique shared environment/rater bias estimates (c) in the Rater Bias model comprised 43% and 47% of the total variance in parent- and teacher-rated ADHD scores, respectively. Because studies have consistently documented that shared environmental factors have little to no influence on ADHD, (c) likely represents rater bias. However, with the addition of the (a) parameter in the better-fitting Psychometric model, the rater bias estimates changed for one of the raters. For parents, the estimates for unique genetic contributions (a) and rater bias (c) were zero and 45%, respectively, suggesting that much of the rating process that is unique to parents is influenced by bias. For teachers, the estimate for unique genetic contributions (a) was 57% and the rater bias (c) estimate decreased to zero, suggesting that teachers display little to no bias in their ratings and are observing an ADHD phenotype that is different from parents, but still valid.

The effect of multiple teachers

In order to determine whether including twins who had been rated by two different teachers would influence our results, we ran the models utilizing the entire sample, which included twins rated by one teacher and twins rated by two teachers. As with the one-teacher analyses, the Psychometric model provided better fit to the data (χ^2 diff(2) = 10.45, $p = .0054$). However, results for the combined sample analyses differed from those for the one-teacher analyses. When examining the parameter estimates from the better-fitting Psychometric model, analyses resulted in a parent bias estimate (c) of .03 and a large estimate for teacher bias (.40), suggesting that teachers, rather than parents, display marked evidence of rater bias (Table 3). Not surprisingly, the measurement error (e) estimate for teachers (.33) was higher than for parents (.13), and higher than when only one teacher rated the twins (.13). Overall, these results suggest that the use of multiple teachers to rate a twin pair likely increases error in the model and inflates rater bias estimates for teachers.

Table 2 Standardized estimates for genetic and environmental contributions

Proportion of total variance in ADHD scores due to:		Rater bias model		Psychometric model	
		Parent	Teacher	Parent	Teacher
Common rater view	genetic effects (A)	.24	.18	.21 (.06–.41)	.17 (.05–.32)
	shared environment (C)	.00	.00	.00	.00
	non-shared environment (E)	.16	.12	.16 (.08–.26)	.13 (.07–.21)
Unique rater view	genetic effects (a)	–	–	.00 (.00–.46)	.57 (.29–.72)
	shared environment/rater bias (c)	.43	.47	.45 (.00–.64)	.00 (.00–.25)
	non-shared environment/measurement error (e)	.17	.23	.18 (.09–.29)	.13 (.05–.22)

Note. Confidence intervals are in parentheses.

Discussion

We utilized multiple rater models to examine parent- and teacher-rated ADHD in selected and control samples of twins in order to explore whether rater disagreement is attributable to rater bias or to the raters observing different but valid ADHD behaviors. We found that a model incorporating genetic contributions unique to a rater provided better fit to the data than a model that did not, suggesting that parents and teachers are observing somewhat different but valid ADHD behaviors in the twins. Furthermore, parameter estimates for the better-fitting Psychometric model showed that the majority of the total variance in ADHD scores (63% for parents and 70% for teachers) was attributable to the rater’s unique view. These results suggest that parents and teachers are observing, to a notable extent, different ADHD phenotypes in these children, which may contribute to rater disagreement. This is not surprising, considering that parents and teachers see the children in different environments where children undergo different tasks.

Our results also suggest that while parents and teachers have different views, parents display signs of bias in their ADHD ratings whereas teachers do not. This result was only true when the twin pairs were rated by one parent and one teacher. Results from analyses that included twin pairs who had been rated by two different teachers showed greater signs

of rater bias and measurement error for teachers, suggesting that using only one teacher to rate a twin pair will significantly increase accuracy.

Overall, these results suggest that teachers may be more reliable reporters of ADHD symptoms than parents. This result is supported by several studies that have shown that parent ratings consistently display signs of sibling contrast effects in twin studies of ADHD (Eaves et al., 1997; Silberg et al., 1996; Thapar et al., 1995), whereas teacher ratings do not. In these studies, parents likely exaggerate differences between their DZ twins because they have no other children with whom to compare. Teachers, however, deal with many students on a daily basis, providing a potential normative group to compare to. This explanation makes sense, as the highly heritable nature of ADHD suggests that DZ twins should be more highly correlated on ADHD symptoms than parent ratings indicate.

Another possibility for the lack of bias in teacher ratings is that ADHD symptoms may be more easily observed by teachers, as the behavioral demands of the classroom may be more stringent than at home. When one examines the DSM-IV ADHD items, many (if not all) of them go directly against what is expected in the classroom. Thus, unless the DSM-IV ADHD criteria are reconsidered, it may be important to place greater emphasis on teacher ratings of ADHD in studies examining this disorder as well as in clinical settings.

Table 3 Standardized estimates for combined sample

Proportion of total variance in ADHD scores due to:		Rater bias model		Psychometric model	
		Parent	Teacher	Parent	Teacher
Common rater view	genetic effects (A)	.23	.15	.18 (.07–.30)	.12 (.05–.20)
	shared environment (C)	.00	.00	.00	.00
	non-shared environment (E)	.22	.15	.23 (.16–.30)	.15 (.11–.20)
Unique rater view	genetic effects (a)	–	–	.43 (.16–.57)	.00 (.00–.18)
	shared environment/rater bias (c)	.33	.38	.03 (.00–.26)	.40 (.23–.49)
	non-shared environment/measurement error (e)	.22	.32	.13 (.07–.21)	.33 (.26–.40)

Note. Confidence intervals are in parentheses.

However, these results must be interpreted with caution: confining our sample only to twins rated by one teacher created limitations on power. This not only decreased the overall sample size significantly, but because the joint twin analyses examine MZ/DZ twins and selected/control twins separately, there were even more limitations on power. Thus, it is possible that the results of the one-teacher analyses may be due to limited sample size and therefore limited power to estimate the parameters with greater accuracy. This is confirmed by the larger confidence intervals for some of the parameters. Studies with a larger sample of twin pairs rated by one parent and one teacher are needed before any firm conclusions can be drawn about rater bias.

In conclusion, this study is the first to utilize twin modeling to examine rater disagreement for ADHD among parents and teachers. Our results suggest that 1) disagreement for ADHD ratings is significantly due to parents and teachers observing different ADHD behaviors, some of which is valid and some of which is due to bias, and 2) parents may be more biased than teachers in their ADHD ratings. These results are tentative and need replication in a larger sample. Although rater disagreement can be a frustration to the researcher, understanding the nature of rater differences may provide valuable information about ADHD as well as other psychiatric disorders.

References

- Aitken, A. C. (1934). Note on selection from a multivariate normal population. *Proc Edinburgh Math Soc*, 55, 42–58.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317–322.
- APA. (1994). *Diagnostic and Statistical Manual of Mental Disorders (4th edition)*. Washington, DC: American Psychiatric Association.
- Barkley, R., & Murphy, K. (1998). *Attention-Deficit Hyperactivity Disorder: A Clinical Workbook* (2nd ed.). New York: Guilford Press.
- Bartels, M., Hudziak, J. J., van den Oord, E., Van Beijsterveldt, C. E. M., Rietveld, M. J. H., & Boomsma, D. I. (2003). Co-occurrence of aggressive behavior and rule-breaking behavior at age 12: multi-rater analyses. *Behavior Genetics*, 33(5), 607–621.
- Costello, E. J., Costello, A. J., Edelbrock, C., Burns, B., Dulcan, M. K., Brent, D., et al. (1988). Psychiatric disorders in pediatric primary care. *Arch Gen Psychiatry*, 45, 1107–1116.
- DeFries, J. C., Filipek, P. A., Fulker, D. F., Olson, R. K., Pennington, B. F., Smith, S. D., et al. (1997). Colorado Learning Disabilities Research Center. *Learning Disabilities*, 8, 7–79.
- DuPaul, G. J., Power, T. P., Anastopoulos, A. D., & Reid, R. (1998). *ADHD Rating Scale – IV*. New York: Guilford.
- Eaves, L., Silberg, J., Meyer, J., & Maes, H. (1997). Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *Journal of Child and Adolescent Psychiatry*, 38, 965–980.
- Gadow, K. D., & Sprafkin, J. (1999). *1999 Revised Norms and Scoring Guide Child Symptom Inventory – 4*. Stony Brook, NY: Checkmate Plus.
- Goodman, R., & Stevenson, J. (1989). A twin study of hyperactivity-II. The aetiological role of genes, family relationships, and perinatal adversity. *Journal of Child Psychology and Psychiatry*, 30, 691–709.
- Hewitt, J. K., Silberg, J., Neale, M. C., Eaves, L., & Erickson, M. T. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics*, 22(3), 293–317.
- Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York: Guilford.
- Lahey, B. B., Applegate, B., McBurnett, K., Biederman, J., Greenhill, L., Hynd, G. W., et al. (1994). DSM-IV field trials for Attention Deficit Hyperactivity Disorder in children and adolescents. *American Journal of Psychiatry*, 151, 1673–1685.
- Neale, M. C. (1997). *Mx: Statistical Modeling* (4th ed.). Richmond, VA: Medical College of Virginia, Department of Psychiatry.
- Neale, M. C., & Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families*. Dordrecht, The Netherlands: Kluwer Academic Publishers B.V.
- Pearson, K. (1902). I: mathematical contributions to the theory of evolution. XI: on the influence of natural selection on the variability and correlation of organs. *Philos Trans R Soc Lond A*, 200, 1–66.
- Piacentini, J. C., Cohen, P., & Cohen, J. (1992). Combining discrepant diagnostic information from multiple sources: are complex algorithms better than simple ones? *J Abnorm Child Psychol*, 20(1), 51–63.
- Sherman, D. K., Iacono, W. G., & McGue, M. K. (1997). Attention-deficit hyperactivity disorder dimensions: A twin study of inattention and impulsivity-hyperactivity. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(6), 745–753.
- Silberg, J., Meyer, J., Rutter, M., Simonoff, E., Hewitt, J. K., Loeber, R., et al. (1996). Genetic and environmental influences on the covariation between hyperactivity and conduct disturbance in juvenile twins. *Journal of Child Psychology and Psychiatry*, 37, 803–816.
- Simonoff, E., Pickles, A., Hervas, A., Silberg, J., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine*, 28, 825–837.
- Simonoff, E., Pickles, A., Meyer, J., Silberg, J., Maes, H., Loeber, R., et al. (1997). The Virginia Twin Study of Adolescent Behavioral Development: influences of age, gender, and impairment on rates of disorder. *Arch Gen Psychiatry*, 54, 801–909.
- Sprafkin, J., Volpe, R. J., Gadow, K. D., Nolan, E. E., & Kelly, K. (2002). A DSM-IV-referenced screening instrument for preschool children: The early childhood inventory – 4. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 604–613.
- Stallings, M. C., Cherny, S. S., Young, S. E., Miles, D. R., Hewitt, J. K., & Fulker, D. F. (1997). The familial aggregation of depressive symptoms, antisocial behavior, and alcohol abuse. *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, 74, 183–191.
- Thapar, A., Harrington, R., Ross, K., & McGuffin, P. (2000). Does the definition of ADHD affect heritability? *Journal of the American Academy of Child and Adolescent Psychiatry*, 39(12), 1528–1536.
- Thapar, A., Hervas, A., & McGuffin, P. (1995). Childhood hyperactivity scores are highly heritable and show sibling competition effects: Twin study evidence. *Behavior Genetics*, 25(6), 537–544.
- Van Der Valk, J. C., van den Oord, E., Verhulst, F. C., & Boomsma, D. I. (2001). Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: Different views or rater bias? *Journal of Child Psychology and Psychiatry*, 42(7), 921–931.
- Willcutt, E. G., Chhabildas, N. A., & Pennington, B. F. (2001). Validity of the DSM-IV subtypes of ADHD. *The ADHD report*, 9, 2–5.
- Willcutt, E. G., Hartung, C. M., Lahey, B. B., Pelham, W. E., & Loney, B. R. (1999). Utility of behavior ratings made by examiners during assessments of children with attention-deficit/hyperactivity disorder. *Journal of Abnormal Child Psychology*, 27, 463–472.