But Fidgety Phil
He won't sit still
He wriggles
And giggles
and then, I declare,
Swings backwards and forwards,
And tilted his chair
See the naughty, restless child,
Growing still more rude and wild,
Till his chair falls over quite.

Freely translated from Heinrich Hoffmann, 1845

Reading committee

Paranimfen

The data collection for this study was supported by .....

Financial support by ..... and CNCR for the publication of this thesis is gratefully acknowledged.

ISBN-10:
ISBN-13:

Printed by:
Cover and lay-out: Pascal Mellema (pascalmellema@hotmail.com)
Cover idea: Eske Derks

VRIJE UNIVERSITEIT

ASSESSMENT AND GENETIC AETIOLOGY OF ATTENTION PROBLEMS, HYPERACTIVITY, AND RELATED DISORDERS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Psychologie en Pedagogiek
op woensdag 29 november 2006 om 13.45 uur
in het auditorium van de universiteit,
De Boelelaan 1105

door

Ester Maria Derks

geboren te Arnhem

promotor:	prof.dr. D.I. Boomsma
copromotoren:	dr. C.V. Dolan
	prof. J.J. Hudziak, MD

Contents

Contents

CHAPTER

1

AIMS AND OUTLINE

In this thesis I will focus on a number of aspects related to attention problems (AP), hyperactivity (HI), and attention deficit hyperactivity disorder (ADHD). These aspects all deal with the analysis of individual differences in AP, HI and ADHD in large cohorts of young twins.The first chapter provides an introduction and background on the genetics of AP, HI, and ADHD in children. Chapters three to twelve are divided into three sections, in accordance with the three main aims of this thesis. Chapter thirteen presents a more detailed description of the sample and the data collection procedures in relation to the effects of attrition and birth cohort on attention problems. Chapter fourteen provides a summary and general discussion of the findings in this thesis.

AIM 1: INVESTIGATION OF SEX DIFFERENCES IN AP, HI, AND ADHD

The first aim of this thesis is to investigate sex differences in AP, HI, and ADHD. In chapter three, I look at the measurement issues and investigate if the relation between latent AP factors and observed item scores for AP, HI, and ADHD is identical in boys and girls. In chapter four, the relation between a quantitative (Child Behavior Checklist) and a categorical (DSM-IV) measure is examined. One of the questions concerns the presence of sex differences in the association between the CBCL-AP score and the DSM-IV diagnosis of ADHD. In chapter five, I address the issue of gender and prevalence, and attempt to answer the question why boys with ADHD are more often referred for treatment than girls with ADHD.

AIM 2: METHODOLOGICAL ISSUES IN TWIN STUDIES ON ATTENTION PROBLEMS

The second aim of this thesis is to look at a number of issues, which are important to the methodology of twin studies. These include potential differences between responders and non-responders in research projects. In chapter six, the impact of non-random attrition is discussed under the assumption that the attrition is related to aspects that are measured at an earlier time-point. Psychopathological data are often highly skewed and kurtotic, because of the absence of psychopathology in most subjects. In chapter seven, I investigate the effect of non-normality of observed phenotypic data on the estimates of the genetic and environmental influences. Estimation of genetic and environmental influences on a phenotypic trait within the twin design is usually based on the assumption that monozygotic (MZ) and dizygotic (DZ) twins share equal amounts of the environment that are relevant to the specific trait under study. In chapter eight, I explore how this "Equal Environment Assumption" can be tested based on multivariate phenotypic data.

AIM 3: STUDYING AETIOLOGICAL INFLUENCES ON INDIVIDUAL DIFFERENCES IN AP, HI, AND ADHD

The third aim of this thesis is the investigation of genetic and environmental influences on individual differences in AP, HI and ADHD. I extent this question by looking at the effects of different informants on the behavior of children and the use of different measures. I will then move toward identifying specific environmental mediators of these traits using the MZ discordant design. In chapter nine, genetic and environmental influences on individual differences in childhood psychopathology in three-year-old children whose behaviour is rated by their mothers and fathers, are examined. In chapter ten, maternal Child Behavior Checklist (CBCL) scores and Teacher Report Form (TRF) scores on AP are analyzed in a multivariate design to examine whether individual differences are explained by the same genetic and environmental influences, or by different influences. The presence of teacher rater bias is investigated by comparing the correlation in twin-pairs in which the two members are rated by the same teacher with the correlation in twins-pairs in which the two members are rated by different teachers. In chapter eleven, multivariate analyses are performed on three maternally rated problem behavior scores: CBCL-AP, Conners ADHD-index, and the DSM-IV diagnosis of ADHD in a selected sample of twins. Chapter twelve, which uses the "MZ discordant twin design" aims to specify environmental influences that make two members of a monozygotic twin-pair (who are genetically identical) different.

CHAPTER

2

GENETICS OF ADHD, HYPERACTIVITY, AND ATTENTION PROBLEMS

This chapter is based on E.M. Derks, J.J. Hudziak, & Boomsma, D.I. Genetics of ADHD, Hyperactivity, and Attention Problems. *In: Handbook of Behavior Genetics, in press.*

OVERVIEW

Attention Deficit Hyperactivity Disorder (ADHD) is characterized by symptoms of inattention, and/or hyperactivity-impulsivity. Inattention symptoms are present when an individual fails to pay attention and has difficulty concentrating. Children or adults who are hyperactive fidget, squirm and move about constantly and can't sit still for any length of time. Impulsivity can be described as acting or speaking too quickly without first thinking of the consequences. Children with ADHD face developmental and social difficulties. As adults, they may face problems related to employment, driving a car, or relationships (Barkley, 2002). As is the case for many other psychiatric disorders, the diagnosis of ADHD is not based on a specific pathological agent, such as a microbe, a toxin, or a genetic mutation, but instead on the collection of signs and symptoms that occur together more frequently than expected by chance (Todd et al., 2005). Genetic studies of psychiatric disorders are complicated by this lack of clear diagnostic tests (Hudziak, 2001). Heritability estimates in epidemiological genetic studies, and the results of gene-finding studies may vary as a consequence of the instrument that is used to assess ADHD, and of other factors such as the specific population that is investigated. In the current chapter we will focus on behavioral measures of ADHD, and not on endophenotypes (i.e., phenotypes that form a link between the biological pathway and the behavioral outcome, for example executive functioning). An excellent overview of endophenotypes for ADHD can be found in Castellanos & Tannock (2002). In this overview, we will first present epidemiological studies on the prevalence of ADHD (section I). Next, the results of studies reporting the heritability of ADHD and related phenotypes will be discussed (section II). We concentrate on variation in these statistics as a result of the specific characteristics of the samples (e.g., age, and sex of the children) and as a result of variation in the assessment methods, and informants. Finally, we give an overview of studies reporting on the agreement between questionnaire data and diagnostic interviews (section III).

PREVALENCE OF ADHD

The current guidelines for the diagnosis of ADHD in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) describe three different subtypes of ADHD: i) ADHD of the inattentive type, which requires the presence of six out of nine symptoms related to inattention; ii) ADHD of the hyperactive/impulsive type, which requires the presence of six out of nine hyperactive/impulsive symptoms, and iii) ADHD of the combined type, which requires the presence of six out of nine inattention symptoms and six out of nine hyperactive/impulsive symptoms (American Psychiatric Association, 1994). Additional criteria are the presence of some hyperactive/impulsive or inattentive symptoms before age 7 years, and impairment from the symptoms in two or more settings. In research settings, the diagnosis of ADHD is not always based on these formal criteria. In some studies, the diagnosis is based on behavior checklists, whose items are summed into a total score. ADHD is then assumed to be present when a child scores above a certain diagnostic cutoff criterion. Diagnoses based on checklists usually do not incorporate additional requirements such as age of onset before age 7 years, or impairment.

Prevalence estimates of ADHD may vary as a result of instrument variance (e.g., DSM diagnoses versus checklists), and as a function of sex and age of the children. We summarize epidemiological studies that report prevalence estimates for ADHD based on DSM criteria in Table 1. These prevalences can be compared with the prevalences based on checklist data which are shown in Table 2. In both Tables, information on the assessment method, and on the age and sex of the children has been included.

The prevalences based on *diagnostic interview studies* varied between 1.5% and 19.0% in boys, and between 0.3 and 8.8% in girls. In both boys and girls, the lowest prevalence was reported in a study that used a three month prevalence instead of the usual 1 year prevalence which may explain the discrepancy with other findings (Costello et al., 2003).

The highest prevalence was reported in a study that did not include impairment criteria (Graetz et al., 2005). Breton et al. (199(1999) also excluded impairment criteria. Excluding the results of these three studies, the prevalences are in the range of 2.4-11% in boys and 1.3-4% in girls. The prevalences based on *checklist data* range between 2.9 and 23.1% in boys and between 1.4-13.6% in girls. Baumgaertel et al. (1995)(1995), who did not show the prevalences by sex, reported a prevalence of 17.8, which is in the upper range for both sexes.

Clearly, higher prevalences are reported when diagnosis is based on questionnaire data compared to clinical diagnoses. How can this discrepancy be explained? Wolraich et al. (1998) showed that the rate of overall ADHD (i.e., irrespective of subtype), based on checklist data in a sample of 698 boys and girls drops from 16.1% to 6.8% when impairment is required for diagnosis. Similarly, in the study of Breton et al. (1999), the prevalence based on parental reports dropped from 5.0 to 4.0% when including impairment criteria. Because impairment criteria are usually included in diagnostic interview studies and not in studies using questionnaire data, it is likely that the higher prevalence in questionnaire data is the result of the exclusion of impairment criteria.

Table 1. Prevalence estimates based on clinical diagnosis in community based samples

Study	N	Method	Any ADHD boys/girls (sex ratio)	Inattentive boys/girls sex ratio)	Hyperactive boys/girls (sex ratio)	Combined boys/girls (sex ratio)	Age
Lavigne et al. (1996)*	1150	DSM-III-R diagnosis by clinician	2.4/1.3 (1.8)	-	-	-	2-5
Breton et al. (1999)**	2400	DSM-III-R clinical interview with child Six month prevalence Impairment not include d	3.3	-	-	-	6-14
	2400	Clinical interview with teacher Six month prevalence Impairment not include d	8.9	-	-	-	6-14
	2400	Clinical interview with parent Six month prevalence Impairment not include d	5.0	-	-	-	6-14
Rohde et al. (1999)	1013	DSM-IV clinical interview with parent and child	5.5/6.1 (.90)	2.0	0.8	3.0	12-14
Cuffe et al. (2001)	490	DSM-III-R clinical interview with adolescent and parent	2.6/0.5 (4.9)	-	-	-	16-22
Graetz et al. (2001)	3597	DSM-IV clinical interview with parent	11.0/4.0 (2.8)	5.1/2.3 (2.2)	2.4/1.4 (1.7)	3.1/0.7 (4.4)	6-17
Costello et al. (2003)	1420	DSM-IV clinical interview with parent Three month prevalence	1.5/0.3 (5)	-	-	-	9-13
Ford et al. (2003)	10438	DSM-IV clinical interview with parent, teacher and self (diagnosis based on judgment by clinician)	3.6/0.9 (4.3)	1.0/0.3 (3.0)	0.3/0.04 (7)	2.3/0.5 (4.6)	5-15

* This is the weighted N which is calculated based on the information provided in the paper. The weigthed prevalence of ADHD is 2%, the number of subjects is 23, so the weighted total number of subjects is 23 /.02=1150

** Breton et al. do not give the prevalences by sex, but do report the odds ratio's for male:female. These are 4.0 in self-reports, 5.1 in teacher reports, and 2.9 in parental reports

Table 1. Prevalence estimates based on clinical diagnosis in community based samples *(Continued)*

Study	N	Method	Any ADHD boys/girls (sex ratio)	Inattentive boys/girls sex ratio)	Hyperactive boys/girls (sex ratio)	Combined boys/girls (sex ratio)	Age
Graetz et al. (2005)	2375	DSM-IV clinical interview with parent Impairment not included	19.0/8.8 (2.2)	8.9/4.4 (2.0)	3.4/1.8 (1.9)	6.7/2.6 (2.6)	6-13
Neuman et al. (2005)	1472	DSM-IV clinical interview with parent	7.4/3.9 (1.9)	4.5/0.6 (7.5)	0.5/1.2 (.4)	2.3/2.1 (1.1)	7-19

* This is the weighted N which is calculated based on the information provided in the paper. The weigthed prevalence of ADHD is 2%, the number of subjects is 23, so the weighted total number of subjects is 23 /.02=1150

** Breton et al. do not give the prevalences by sex, but do report the odds ratio's for male:female. These are 4.0 in self-reports, 5.1 in teacher reports, and 2.9 in parental reports

Table 2. Prevalence estimates based on behavioral checklist data in community based samples

Study	N	Method	Any ADHD boys/girls (sex ratio)	Inattentive boys/girls (sex ratio)	Hyperactive boys/girls (sex ratio)	Combined boys/girls (sex ratio)	Age
Szatmari et al. (1989)	1486	DSM-III-R rating scale by parent, teacher, and self. Prevalences based on hierarchical log-linear models	10.1/3.3 (3.1)				4-11
	1236	DSM-III-R rating scale by parent, teacher, and self. Prevalences based on hierarchical log-linear models	7.3/3.4 (2.1)	-	-	-	12-16
Baumgaertel et al.(1995)	1077	DSM-IV rating scale by teacher	17.8	9.0	3.9	4.8	5-12
Wolraich et al. (1996)	8258	DSM-IV rating scale by teacher	16.2/6.1 (2.7)	7.2/3.5 (2.1)	3.8/0.9 (4.2)	5.3/1.6 (3.3)	Kindergarten through 5th grade
Nolan et al. (2001)	413	DSM-IV rating scale by teacher	21.5/13.6 (1.6)	3.8/4.0 (.95)	7.6/5.1 (1.5)	10.1/4.6 (2.2)	3-5
	1520	DSM-IV rating scale by teacher	23.1/8.2 (2.8)	14.4/6.0 (2.4)	3.4/1.1 (3.1)	5.3/1.1 (4.8)	5-12
	1073	DSM-IV rating scale by teacher	20.1/8.8 (2.3)	14.5/8.0 (1.8)	1.6/0.0 (incalculable)	4.0/0.8 (5.0)	12-18
Larsson et al. (2004)	2063	DSM-III-R rating scale by parent	4.7	-	-	-	8-9
	2055	DSM-III-R rating scale by parent	3.1	-	-	-	13-14
Levy et al. (2005)	1550	DSM-IV rating scale by mother	-	1/4.3 (2.3)	3.1/1.7 (1.8)	5.8/2.0 (2.9)	4-12
Cuffe et al. (2005)	6933	Strengths and Difficulties Questionnaire	3.1/1.4 (2.2)	-	-	-	4-8
	7431	Strengths and Difficulties Questionnaire	6.3/2.1 (3.0)	-	-	-	9-13
	5636	Strengths and Difficulties Questionnaire	2.9/1.8 (1.6)	-	-	-	14-17

In Tables 1 and 2, higher prevalences for ADHD are reported in boys than in girls. The mean sex ratio's were calculated by taking the average of the sex ratio's across studies. For overall ADHD, the ratio of boys:girls ranges from 0.9:1 to 5:1 with a mean sex ratio of about 2.5:1. The sex ratio is lowest in young children (3-5 years; mean sex-ratio is 1.7:1), and highest in older children (5-13 years; mean sex-ratio is about 3:1). In adolescents (13-17 years), the sex-ratio is about 2.5:1. The sex ratio's do not vary much by subtype. The sex ratio's are 2.5:1., 2.5:1, and 3.5:1 for the inattentive type, the hyperactive-impulsive type, and the combined type, respectively. The male:female ratio is not very high in epidemiological studies (about 3:1), but is clearly higher (about 9:1) in clinical settings (Gaub & Carlson, 1997).

In two studies, the prevalence of ADHD was estimated separately in three age groups (Nolan et al., 2001; Cuffe et al., 2005). Both studies show a relatively low prevalence of ADHD in young children, an increased prevalence in older children, and again a relatively low prevalence in adolescents. A recent epidemiological study in adults showed that ADHD may be common in adulthood. Broad screening DSM-IV criteria (symptom occurred sometimes or often) identified 16.4% of a population of 966 adults as having ADHD, while 2.9% of the adults met narrow screening criteria (symptom occurred often) (Faraone & Biederman, 2005).

Genetic epidemiological studies on ADHD in children

Many studies report the heritability of ADHD from a comparison of the covariance structure in monozygotic (MZ) and dizygotic (DZ) twins. In these studies, variation in the vulnerability for ADHD is decomposed into genetic and environmental components. The decomposition of variance takes place by comparing the similarity (covariance or correlation) between MZ twins, who are nearly always genetically identical, and DZ twins, who on average share half of their segregating alleles. MZ twins share all additive genetic and non-additive genetic variance. DZ twins on average share half of the additive genetic and one quarter of the non-additive genetic variance (Plomin et al., 2001). The environmental decomposition of the phenotypic variance is into shared environmental variance and non-shared, or specific, environmental variance. The environmental effects shared in common by two members of a twin pair (C) are by definition perfectly correlated in both monozygotic and dizygotic twins. The non-shared environmental effects (E) are by definition uncorrelated in twin pairs. A first estimate of additive genetic heritability based on twin data is obtained from comparing MZ and DZ correlations: $a^2 = 2(r_{MZ} - r_{DZ})$. The importance of non-additive genetic influence is obtained from: $d^2 = 4(r_{DZ} - r_{MZ})$ and of shared environmental factors $c^2 = 2r_{DZ} - r_{MZ}$. Finally, the estimate of the non-shared environmental component is obtained from $e^2 = 1 - r_{MZ}$. In the classic twin design, one cannot estimate D and C simultaneously and usually the choice for an ADE or ACE model is based on the pattern of MZ and DZ twin correlations. Parameters a^2, c^2, d^2, and e^2 are then obtained with e.g. Maximum likelihood estimation using software packages as Mx (Neale et al., 2003) or Mplus (Muthén & Muthén, 2000).

Papers reporting on the heritability of ADHD find large genetic influences, irrespective of the choice of instrument, informant, or sex and age of the child. Another general finding is the non-significant influence of the shared environment. We summarize these results by measurements of: i) ADHD symptoms (i.e., instrument includes both hyperactivity-impulsivity and attention problem symptoms (Table 3); ii) hyperactivity (Table 4); and iii) attention problems (Table 5). In the tables, we included information on the instrument that was used to assess ADHD. It should be noted that the majority of the studies used symptom counts rather than categorical diagnosis. If a research group published more than one paper based on the same sample, we included only the study with

the largest sample size. The broad-band heritability of ADHD ranges between 35% and 89%. For hyperacti-vity, the broad-band heritability ranges between 42 and 100%. Finally, for attention pro-blems, the broad-band heritability ranges between 39 and 81%.

Longitudinal studies show that symptom ratings of attention problems are stable between ages 7 and 12 (Rietveld et al., 2004). The same is true for symptom ratings of ADHD between 8 and 13 years of age (Larsson et al., 2004). These two studies report remarkably similar correlations of about .5 for five year test-retest correlations. Likewise, both studies report that the stability of symptom ratings of attention problems is mainly explained by additive genetic effects, but that the genetic effects are far from perfectly stable. Only a subset of the genes that operate at one age does so at a later age.

Sex differences in genetic influences on ADHD

When examining the genetic architecture of a trait, two different kinds of sex differences can be distinguished. *Quantitative sex differences* reflect sex differences in the magnitude of the genetic influences: do genes explain the same or different amounts of variation in boys and girls? *Qualitative sex differences* reflect differences in the specific genes that are expressed in boys and girls. Below, we discuss quantitative and qualitative sex differences in ADHD.

Thirteen of the studies reported in Tables 3-5 tested for quantitative sex differences in ADHD (see Table 3-5). Seven of these studies reported the absence of significant sex differences. In the remaining six studies, the presence of sex differences varied by informant and age. The effect sizes of the statistically significant sex differences were small and the pattern of sex differences was inconsistent over studies. In some studies heritability was higher in boys while in other studies heritability was higher in girls. The small effect sizes and the inconsistent pattern of results support the conclusion that the magnitudes of the etiological factors influencing variation in ADHD do not vary much as a function of the child's sex.

Nine studies investigated if different genes are expressed in boys and girls. Eight studies did not find qualitative sex differences. One study reported on different genes in boys and girls, but only for twins who were rated by the same teacher and not for twins rated by parents or different teachers (Saudino et al., 2005). Future studies should reveal if this finding of qualitative sex differences in teacher ratings can be replicated.

Table 3. Heritability estimates based on epidemiological studies of ADHD

Study	Age	N pairs	Sample and Assessment instrument	Rater	Heritability (A+D)	Quantitative sex difference in heritability	Contrast or dominance effect
Eaves et al. (1997)	8-16	1412	Virginia Twin study of Adolescent Behavioral Development Diagnostic interview	Mother and Father	71(boys)/74 (girls) (Mother) 78 (boys)/55 (girls) (Father)	No (mother) Yes (father)	AE_S model fits better than ADE model
	8-16	1412	Virginia Twin study of Adolescent Behavioral Development Questionnaire	Mother, Father, and Teacher	62 (boys) /54 (girls) (Teacher) 75 (boys) /63 (girls) (Mother) 82 (boys) /76 (girls) (Father)	No	Mother and Father: AE_S model fits better than ADE model Teacher: AE model
Sherman et al. (1997b)	11-12	288	Minnesota twin family study Mother: Diagnostic interview Teacher: Rating For m	Mother and Teacher	89 (Mother) 73 (Teacher)	-	-
Coolidge et al. (2000)	Mean=9	112	Coolidge Personality and Neuropsychological Inventory	Parent	82	-	AE model
Thapar et al. (2000)	5-17	1321*	Greater Manchester Twin Register DuPaul rating scale (parent and teacher), and Rutter A scale (parent)	Parent and teacher	78 (DuPaul, parent) 84 (Rutter, parent) 77 (DuPaul, teacher)	No	Parent, DuPaul): ADE (var MZ= varDZ) Parent, Rutter: AE_S (var DZ > var MZ) Teacher: AE model
Young et al. (2000)	12-18	334	Colorado Drug Research Center. Diagnostic interview	Self	49	-	AE model
Burt et al. (2001)	10-12	753	Minnesota Twin Family Study Diagnostic Interview	Symptom present when mother or self report it	57	No	-
Price et al. (2001)	2, 3, and 4	3118 (2) 2796(3) 2452(4)	Twins Early Development Study Conners Rating scale	Parent	83 (boys)/79(girls) (age 2) 81 (age 3) 79 (age 4)	Yes (age 2) No (ages 3,4)	AE_S model fits better than ADE model (ages 2, 3, and 4)
Larsson et al. (2004)	8-9 and 13-14	1106 (8-9) 1063 (13-14)	Young Twins Study DSM-III-R based questionnaire	Parent	35(boys)/68(girls (age 8-9) 74(boys)/61(girls) (age 13-14)	Yes	-

* The authors report in the abstract that data from 2082 pairs are included, but 162 pairs with unknown zygosity and 599 opposite sex pairs are excluded from statistical analyses

AE_S model includes additive genetic effects, non-shared environmental effects and contrast effects; ADE model includes additive genetic effects, non-additive genetic effects, and non-shared environmental effects; var=variance

Table 3. Heritability estimates based on epidemiological studies of ADHD *(Continued)*

Study	Age	N pairs	Sample and Assessment instrument	Rater	Heritability (A+D)	Quantitative sex difference in heritability	Contrast or dominance effect
Dick et al. (2005)	14	631	FinnTwin12 Diagnostic interview	Self	70	-	-
Kuntsi et al. (2005)	Mean=8	3541	Twins Early Development Study Conners Rating scale	Parent	72	No	-
Hudziak et al. (2005)	7	1585	Netherlands Twin Register Conners Rating Scale	Mother	78	No	ADE model (var MZ= var DZ)

* The authors report in the abstract that data from 2082 pairs are included, but 162 pairs with unknown zygosity and 599 opposite sex pairs are excluded from statistical analyses
AE_s model includes additive genetic effects, non-shared environmental effects and contrast effects; ADE model includes additive genetic effects, non-additive genetic effects, and non-shared environmental effects; var=variance

Table 4. Heritability estimates based on epidemiological studies of hyperactivity

Study	Age	N pairs	Sample and Assessment instrument	Rater	Heritability (A+D)	Quantitative sex difference in heritability	Contrast or dominance effect
Goodman & Stevenson (1989)*	13	213	Medical Research Counsel and Inner London Education Authority Rutter behavior questionnaire	Mother, Father and Teacher	>100% (mother) 42% (father) 42% (teacher)	-	-
Thapar et al. (1995)	8-16	376	Cardiff Births Survey Rutter A scale	Mother	88	-	AE_S model (var DZ>var MZ)
Sherman et al. (1997a)	11-12	288	Minnesota twin family study. Teacher: Rating form. Mother: Diagnostic interview	Teacher and Mother	69 (teacher) 91 (mother)	-	-
Simonoff et al. (1998)	8-16	1044	Virginia Twin study of Adolescent Behavioral Development Questionnaire	Mother and Teacher	75 (mother same teacher) 69 (mother different teacher) 52 (teacher)	No	Mother: AE_S model fits better than ADE model Teacher: AE model
Kuntsi & Stevenson (2001)	7-11	268	Twins from primary schools in Southern England Conners Rating Scale	Parent and Teacher	71 (Parent) 57 (teacher)	-	Parent: AE_S model fits better than ADE model Teacher: AE model

* The heritability is more than 100% while the authors calculated A, C, and E based on the MZ and DZ correlations, whereas the DZ correlations were lower than half the MZ correlations
AE_s model includes additive genetic effects, non-shared environmental effects and contrast effects; ADE model includes additive genetic effects, non-additive genetic effects, and non-shared environmental effects; var=variance

Table 4. Heritability estimates based on epidemiological studies of hyperactivity *(Continued)*

Study	Age	N pairs	Sample and Assessment instrument	Rater	Heritability (A+D)	Quantitative sex difference in heritability	Contrast or dominance effect
Martin et al. (2002)	5-16	682 (Parent) 443 (Teacher)	Cardiff Birth Survey Conners Rating Scale (CRS), and Strengths and Difficulties Questionnaire (SDQ)	Parent and Teacher	74 (Parent, CRS) 72 (Parent, SDQ) 80 (Teacher, CRS) 81 (Teacher, SDQ)	-	Parent (CRS and SDQ): ADE model fits better than AE_S model Teacher (CRS and SDQ): AE model
Derks et al. (2004)	3	6250 (Father) 9445 (Mother)	Netherlands Twin Register Child Behavior Checklist 2/3	Father and Mother	66 (Father) 70 (Mother)	No	ADE model (var MZ=var DZ)
Vierikko et al. (2004)	11-12	1636	FinnTwin12 Multidimensional Peer Nomination Inventory	Parent and Teacher	78 (boys, parent) 81 (girls, parent) 49 (boys, teacher) 55 (girls, teacher)	No (parent) Yes (teacher)	Parent (girls): AE_S model (but var MZ=var DZ) Parent (boys): AE model Teacher: AE model
Saudino et al. (2005)	7	3714	Twins Early Development Study Strengths and Difficulties Questionnaire	Parent and Teacher	77 (parent, boys) 75 (parent, girls) 74 (same teacher, boys) 76 (same teacher, girls) 66 (different teacher, boys) 55 (different teacher, girls)	No	-

* The heritability is more than 100% while the authors calculated A, C, and E based on the MZ and DZ correlations, whereas the DZ correlations were lower than half the MZ correlations

AE_s model includes additive genetic effects, non-shared environmental effects and contrast effects; ADE model includes additive genetic effects, non-additive genetic effects, and non-shared environmental effects; var=variance

Table 5. Heritability estimates based on epidemiological studies of attention problems

Study	Age	N pairs	Sample and Assessment instrument	Rater	Heritability (A+D)	Quantitative sex difference in heritability	Contrast or dominance effect
Edelbrock et al. (1995)	7-15	181	Western Reserve Twin Project Child Behavior Checklist	Parent	66	-	
Gjone et al. (1996a)	5-9 12-15	390 (5-9) 526 (12-15)	Norwegian Medical Birth Registry Child Behavior Checklist	Parent	73-76 (age 5-9) 75-79 (age 12-15)	-	-
Sherman et al. (1997a)	11-12	288	Minnesota Twin Family Study. Teacher: Rating form Mother: Diagnostic interview	Mother and Teacher	69 (Mother) 39 (Teacher)	-	-
Hudziak et al. (2000)	8-12	492	Missouri Twin Study Child Behavior Checklist	Parent	60-76	-	No contrast effect, D was not tested
Schmitz & Mrazek (2001)	4-11	207	Colorado Department of Health Statistics Child Behavior Check-list	Mother	54	-	-
Rietveld et al. (2003b)*	7 10 12	3373 2485 1305	Netherlands Twin Register Child Behavior Checklist	Mother	71 (age 7) 70 (boys, age 10) 71 (girls, age 10) 69 (boys, age 12) 73 (girls, age 12)	No (7) Yes(10) Yes (12)	An ADE and an AEs model both provide a good fit to the data
Beijsterveldt van et al. (2004)	5	7679 (mother) 6999 (father)	Netherlands Twin Register Devereux Child Behavior rating scale	Mother and Father	79 (mother boys) 81 (mother girls) 76 (father boys and girls)	Yes (mother) No (father)	AE_S model (var DZ>var MZ)

* Both the ADE and the AEs model provided a good fit at some ages. The heritability estimates are based on the ADE model because this model provided a good fit at all ages
AE_s model includes additive genetic effects, non-shared environmental effects and contrast effects; ADE model includes additive genetic effects, non-additive genetic effects, and non-shared environmental effects

Informant differences

The heritabilities for ADHD rated by father and mother appear to be similar in most studies (Derks et al., 2004; Beijsterveldt van et al., 2004; Eaves et al., 1997), but not in others (Goodman & Stevenson, 1989). Heritabilities for teacher ratings range between 39 and 81%, and are usually lower than the heritabilities based on parental ratings in the same sample (Eaves et al., 1997; Kuntsi & Stevenson, 2001; Simonoff et al., 1998; Vierikko et al., 2004), but see (Martin et al., 2002).

A complexity encountered when teacher ratings are analyzed is that both members of a twin-pair may be rated by the same teacher or by different teachers. Twin correlations are usually higher in children rated by the same teacher than in children rated by different teachers (Saudino et al., 2005; Vierikko et al., 2004; Towers et al., 2000; Simonoff et al., 1998) but not in Sherman et al. (1997a). Simonoff et al. (1998) developed two different models to explore this finding. One model was based on the assumption that teachers have difficulty distinguishing the two children ("twin confusion model"). The other model was based on the assumption that ratings by the same teacher are correlated because i) raters have their own subjective view on which behaviors are appropriate and which are not, or ii) raters influence the behavior of the child because of the rater's own personality characteristics ("correlated errors model"). Although Simonoff et al. (1998) were not able to differentiate between these two models, Derks et al. (in press) reported a better fit of the correlated errors model in a large sample of Dutch twins rated by their teacher.

Selected samples (DeFries-Fulker regression)

Several twin studies have based heritability estimates for ADHD on data from subjects who were selected on a high vulnerability for ADHD. In some of these studies, the subjects with a high vulnerability were selected based on a clinical diagnosis of ADHD, in others they obtained a high behavior checklist score. DeFries & Fulker (1985) developed a multiple regression model that is especially appropriate for the analysis of data in twin pairs in which one member of a pair has been selected because of a deviant score. The rationale of this method is based on the fact that when probands are selected based on high scores on a heritable trait, MZ cotwins are expected to obtain higher scores on the trait than DZ cotwins because of a lower degree of regression to the mean. In the regression model, the cotwin's score is predicted from a proband's score (P) and the coefficient of relationship (R). The coefficient of relationship equals .5 and 1 in DZ and MZ twins, respectively. The basic regression model is as follows: $C = B_1 P + B_2 R + A$, where C is a cotwin's predicted score; B_1 is the partial regression of the cotwin's score on the proband's score; B_2 is the partial regression of the cotwin's score on the coefficient of relationship; and A is the regression constant. B_1 is a measure of twin resemblance that is independent of zygosity. A significant regression coefficient B_2 indicates that being a member of the affected group is heritable. The extreme group heritability (h_g^2), equals: $h_g^2 = B_2$ / (mean score proband - mean score cotwin). After establishing the heritability of the condition by testing the significance of B_2, direct estimates of h^2 (the extent to which individual differences in the unselected population are heritable) and c^2 (the extent to which individual differences in the unselected population are explained by shared environmental factors) can be obtained by fitting the following extended regression model: $C=B_3P + B_4R + B_5PR + A$, where PR is the product of the proband's score and the coefficient of relationship R. B_5 is a direct estimate of h^2, while B_3 is a direct estimate of c^2. DeFries & Fulker (1985) note that if affected individuals represent the lower end of a normal distribution of individual differences, the estimate of h^2 (heritability of the trait in the unselected sample) should be similar to the estimate of h_g^2 (heritability of extreme group membership).

The DeFries-Fulker regression model has been used to estimate h_g^2 and h^2 in a number of studies (Gillis et al., 1992; Stevenson, 1992; Rhee et al., 1999). Gillis et al. (1992) studied the heritability of ADHD in a sample of 74 twin-pairs in which at least one of the

twin members was diagnosed with ADHD. They report an estimate of .98 (±.26) for h_g^2. This is in agreement with an estimate of .81 (±.51) for h_g^2 based on hyperactivity scores in a sample of 196 13 year old twin-pairs (Stevenson, 1992), although this latter estimate did not reach significance.

A number of studies showed that h_g^2 does not vary as a function of the diagnostic cutoff score that is used for assessing ADHD (Willcutt et al., 2000; Price et al., 2001; Levy et al., 1997). Gjone et al. (1996b) also report an absence of change in group heritability with increasing severity, but a slight tendency toward decreased heritability in the more severely affected groups. This suggests that the extreme group heritability does not vary as a function of the diagnostic cutoff score, although there may be a somewhat lower heritability of ADHD at the extreme of the distribution.

An interesting application of DeFries-Fulker regression was shown in Willcutt et al. (2000) who studied ADHD in 373 8 to 18 year old twin-pairs. They investigated if h_g^2 of inattention varies as a function of the level of hyperactivity/impulsivity, and vice versa, if h_g^2 of hyperactivity/impulsivity varies as a function of the level of inattention. The etiology of extreme inattention was similar whether the proband exhibited low or high levels of hyperactivity/impulsivity. In contrast, the heritability of extreme hyperactivity/impulsivity was high in individuals who show high levels of inattention while it was low and non-significant in individuals with low levels of inattention.

The relation between questionnaire data and diagnostic interviews

Derks et al. (2006) reviewed studies that investigated the relation between behavior checklist scores on attention problems and the clinical diagnosis for ADHD and reported on the positive and negative predictive power, sensitivity, and specificity. Many of these studies used the attention problem scale of the Child Behavior Checklist to predict ADHD. Despite its name, the scale also contains items related to hyperactivity-impulsivity. Positive predictive power (PPP) refers to the proportion of children with a high checklist score who obtain a positive DSM diagnosis (i.e. affected), and negative predictive power (NPP) refers to the proportion of children with a low checklist score who obtain a negative DSM diagnosis (i.e., unaffected). Sensitivity and specificity refer to the proportion of children with a positive DSM diagnosis, who score high on the checklist, and the proportion of children with a negative DSM diagnosis, who score low on the checklist, respectively. Table 6 summarizes the results of the studies that used these Diagnostic Efficiency Measures (DES). A negative feature of the DES is their dependence on the baseline prevalence of the disorder. Therefore, the baseline prevalence was also included in Table 6. On the basis of the results, we can conclude that the association between behavior checklist scores and clinical diagnoses for ADHD is strong. However, in population based studies, a low score on the behavior checklist is highly predictive of the absence of ADHD while a high score is less predictive of ADHD. Derks et al. (2006) further showed that a boy with a high CBCL-score has a higher chance of obtaining a positive diagnosis for ADHD than a girl with a high CBCL-score. In other words, questionnaire scores better predict clinical diagnosis in boys than girls.

Table 6. Diagnostic Efficiency Statistics of studies that examined the association between behavior checklist scores and ADHD

Study	Sample	N Boys/Girls	Cutpoint	Prevalence (%)	PPP	NPP	SE	SP
Gould et al. (1993)	NR	157	T > 65	23	.36	.96	.46	.95
Chen et al. (1994)	SR	111/108	T 65	16/8	1.00 (boys) .67 (girls)	.86 (boys) .93 (girls)	.17 (boys) .22 (girls)	1.00 (boys) .99 (girls)
Eiraldi et al. (2000)	R	192/50	T 65	83	.93	.37	.78	.69

R=clinically referred sample, NR=non-referred sample, SR=siblings of referred children
PPP=Positive Predictive Power, NPP=Negative Predictive Power, SE=Sensitivity, SP=Specificity

Table 6. Diagnostic Efficiency Statistics of studies that examined the association between behavior checklist scores and ADHD *(Continued)*

STUDY	SAMPLE	N BOYS/GIRLS	CUTPOINT	PREVALENCE (%)	PPP	NPP	SE	SP
Lengua et al. (2001)	R	203	Based on regression	29	.50	.71	.02	.99
Sprafkin et al. (2002)	R	247/0	T 60	71	.78	.83	.97	.33
Hudziak et al. (2004)	SR	101/82	T 65	36	.97	.76	.47	.99
Derks et al. (in press)	NR	192/216	Longitudinal	14/12	.59/.36	.96/.97	.74/.80	.92/.81

R=clinically referred sample, NR=non-referred sample, SR=siblings of referred children
PPP=Positive Predictive Power, NPP=Negative Predictive Power, SE=Sensitivity, SP=Specificity

In the field of behavioral genetics, the focus of interest is not only on the genetic and environmental influences on the variance of a trait, but also on the genetic and environmental influences on the covariance of two traits. Future studies should investigate the aetiology of the covariance between behavior checklist scores and DSM-IV diagnoses of ADHD. An important issue that needs to be addressed is the overlap of the genetic factors that explain variation in different measures of ADHD.

CURRENT TOPICS

In the previous sections we gave an overview of the results of epidemiological studies on ADHD. A few general findings emerged, among which a higher prevalence of ADHD in boys than girls, and a high heritability of ADHD in children irrespective of sex, age, or informant. In section IV-IX, we discuss current topics in the research field of ADHD. Section IV addresses the question if measurement instruments assess ADHD equally well in boys and girls. Section V discusses the controversy between studies claiming the presence of contrast effects versus non-additive genetic effects on individual differences in ADHD. In section VI we report on the results of genetic analyses in which the ratings from multiple informants are analyzed simultaneously. Sections VII and VIII show two applications of latent class analyses: examination of genetic heterogeneity of the ADHD subtypes, and investigation of the categorical versus continuous distribution of the liability for ADHD. Finally, in section IX, we provide a brief overview of the results obtained in gene-finding studies on ADHD.

MEASUREMENT INVARIANCE WITH RESPECT TO SEX

The prevalence of ADHD is about 2.5 times higher in boys than girls, and there are sex differences in the association between checklist scores and clinical diagnoses. Heritability seems not to vary much as a function of the child's sex, and only one out of nine studies suggests that different genes are expressed in boys and girls.

Before any sex differences in ADHD can be interpreted, we should first establish if the measurement instrument is not biased with respect to sex. Stated differently, the instrument should measure the same construct, i.e., latent variable of interest, in boys and girls (Mellenbergh, 1989; Meredith, 1993). If this is the case then we expect the observed score (i.e., the score obtained on the measurement instrument) of a person to depend on that person's score on the latent construct, but not on that person's sex. If this is not the case, a boy and a girl with identical levels of problem behavior may obtain systematically (i.e., regardless of measurement error) different scores on the instrument. This is undesirable because obviously we wish our measurements to reflect accurate and interpretable differences between cases in different groups. If the measurement instrument is not biased with respect to sex, we say that it is measurement invariant (MI) with respect to sex.

The criteria of MI are empirically testable in the common factor model (Meredith, 1993). Factor analysis may be viewed as a regression model in which observed variables (e.g., item scores) are regressed on a latent variable or common factor. In terms of this regression, the MI criteria are: 1) equality of regression coefficients (i.e., factor loadings) over groups; 2) equality of item intercepts over groups (i.e., differences in item means can only be the result of differences in factor means), and 3) equality of residual variances (i.e., variance in the observed variables, not explained by the common factor) over groups. When satisfied, these restrictions ensure that any group differences in the mean and variance of the observed variables are due to group differences in the mean and variance of the latent factor.

In a sample of 800 boys and 851 girls rated by their teacher, Derks et al. (in revision) established measurement invariance with respect to sex for the Cognitive problems-inattention scale, the Hyperactive scale, and the ADHD-index of the Conners Teacher Rating Scale-Revised. This implies that teacher ratings on ADHD are not biased as a result of the child's sex. Although future studies should show if measurement invariance is also tenable for parental ratings on ADHD, the results in teacher ratings suggest that sex differences in the prevalence of ADHD, and on the predictive value of questionnaire scores are not the result of measurement bias.

Genetic dominance or rater bias/sibling interaction

When reviewing the literature on ADHD, it is remarkable that many studies report very low DZ correlations for parental ratings but not for teacher ratings on ADHD. Low DZ correlations can either be explained by the presence of non-additive genetic effects (Lynch & Walsh, 1998) or by social interaction. The effects of social interaction among siblings were discussed by Eaves (1976) and others (Carey, 1986; Boomsma, 2005). Social interactions between siblings may create an additional source of variance and can either be cooperative (imitation) or competitive (contrast). Cooperation implies that behavior in one sibling leads to similar behavior in the other siblings. In the case of competition, the behavior in one child leads to the opposite behavior in the other child.

In the classical twin design, cooperation, or positive interaction, leads to increased twin correlations for both monozygotic (MZ) and dizygotic (DZ) twins. The relative increase is larger for DZ than for MZ correlations, and the pattern of correlations thus resembles the pattern which is seen if a trait is influenced by the shared environment. Negative sibling interaction, or competition, will result in MZ correlations which are more than twice as high as DZ correlations, a pattern also seen in the presence of non-additive genetic effects.

In data obtained from parental ratings on the behavior of their children, the effects of cooperation and competition may be mimicked (Simonoff et al., 1998). When parents are asked to evaluate and report upon their children's phenotype, they may compare the behavior of siblings. Parents may either stress similarities or differences between children, resulting in an apparent cooperation or competition effect. The presence of a contrast effect, either caused by social interaction or rater bias, is indicated by differences in MZ and DZ variances. If there is a contrast effect the variances of MZ and DZ twins are both decreased, and this effect is greatest on the MZ variance. Contrast and non-additive genetic effects can theoretically be distinguished by making use of the fact that contrast effects lead to differences in variances in MZ and DZ twins while non-additive genetic effects do not. However, Rietveld et al. (2003a) showed that the statistical power to separate these effects is low in the classical twin design.

In Tables 3, 4, and 5, we included information on the influence of non-additive genetic effects and contrast effects on individual differences in ADHD. In the fourteen studies testing for the presence of these effects, a consistent finding was the absence of non-additive genetics and contrast effects in teacher ratings. In parental ratings, nine studies

reported significant contrast effects. However, one of these studies did not report larger variances in DZ than MZ twins, and the presence of non-additive genetic effects was not considered (Vierikko et al., 2004). Another study reported significant contrast effects on the Rutter scale, but significant non-additive genetic effects on the DuPaul rating scale (Thapar et al., 2000). The authors argue that rater contrast effects may be more pronounced for some scales, as a result of differences in the number of items or in the format of the questionnaires. The influence of non-additive genetic effects was also reported in two other studies on hyperactivity. Furthermore, Rietveld et al. (2003b) reported that a model with non-additive genetic effects and a model with contrast effects both provided a good fit to the data. Finally, two studies found no significant influences of either contrast or non-additive genetic effects. Teacher ratings do not indicate the presence of either one of these influences, suggesting that rater bias rather than genetic dominance plays a role in parental ratings. However, this is contradicted by the non-significant variance differences in MZ and DZ twins in some studies. So far, the results on the presence of non-additive genetic effects or contrast effects in parental ratings on ADHD are inconclusive. The issue may be resolved by including ratings from other family members which increases the statistical power to detect genetic dominance.

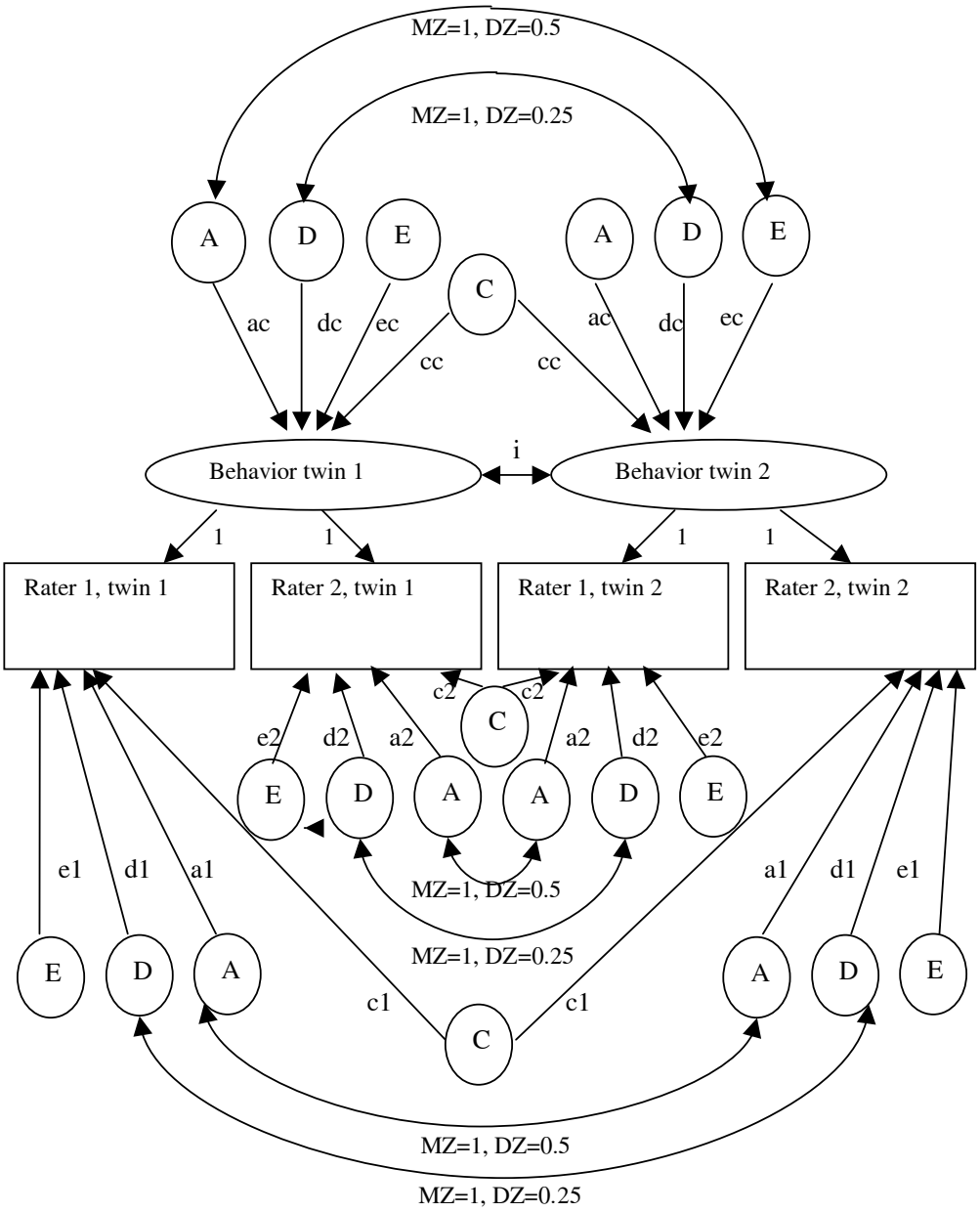
Multiple informants

When investigating genetic and environmental influences on individual differences in problem behavior, we should acknowledge the fact that ratings of problem behavior may be influenced by the rater's personal values and by the unique settings in which the rater and child co-exist. Agreement between raters shows that some aspects of the behavior can be reliably assessed across settings and by different informants. Disagreement may reflect the fact that different raters assess unique aspects of the behavior, which are apparent in a particular set of circumstances, but not in others. For example, a child's inability to concentrate or sit still may be obvious in the classroom setting, but less evident in other settings, where sustained attention is less important (e.g., at play or at home with family members). For CBCL-AP scores, paternal and maternal ratings correlate .73 while parent and teacher correlations show a lower correlation of .44 (Achenbach & Rescorla, 2001).

Different models for twins rated by multiple informants have been developed. In this chapter, we will restrict the discussion to the psychometric model (Hewitt et al., 1992; Neale & Cardon, 1992).

In the Psychometric Model (see Figure 1), the ratings of the child's behavior are allowed to be influenced by aspects of the child's behavior that are perceived by both raters (common factor), and by aspects of the child's behavior that are perceived uniquely by each rater (rater-specific factors). Unique perceptions could arise if the child behaves differentially towards his or her parents, or if the parents observe the child in different situations. The common and unique aspects are both allowed to be influenced by genetic and environmental factors.

Figure 1. Rater Model



Note: The illustrated model is a psychometric model. Both twins are rated by two informants (rater 1 and rater 2). Variation in behavior is explained by common A, C or D, and E (shown in the upper part of the figure), and rater-specific A, C or D, and E (shown in the lower part of the figure). A=additive genetic factor; D=dominant genetic factor; C=shared environmental factor; E=non-shared environmental factor; ac=additive genetic common; dc=dominant genetic common; ec=non-shared environment common; cc=shared environment common; a1=additive genetic rater 1; d1=dominant genetic rater 1; e1=non-shared environment rater 1; c1=shared environment rater 1; a2=additive genetic rater 2; d2=dominant genetic rater 2; e2=non-shared environment rater 2; c2=shared environment rater 2, i=social interaction path

Maternal and *paternal* ratings on overactive behavior in three-year-olds correlate between .66-.68 in boys, girls, and opposite-sex twins. Bivariate analyses showed that 68% of the variance is explained by a factor that is stable across informant (Derks et al., 2004). The remaining variance is explained by rater-specific factors. The heritability of the common factor is high (72%). In addition, genes explain more than half of the variation of the rater-specific factors (55% for fathers and 67% for mothers). The fact that variation in the rater-specific factors is not completely explained by environmental factors, implies that disagreement between parents is not only the result from rater-specific views (i.e., measurement error). In contrary, paternal and maternal ratings are influenced by aspects of the child's behavior that are uniquely perceived by each parent.

To determine how much of the variation in *parent* and *teacher* ratings is due to rating similar versus situation specific components of behavior, some investigators employed bivariate model fitting analyses, which revealed that maternal and teacher ratings partly reflect a common latent phenotype (Martin et al., 2002; Simonoff et al., 1998; Derks et al., in press). In Martin et al., 42% of the variation in the Strengths and Difficulties Questionnaire (SDQ) is explained by a factor that is common to parent and teacher ratings, the heritability of this factor is 90%. The heritability of the rater specific factors is 22% in parent ratings and 65% in teacher ratings. The authors also obtained parental and teacher Conners Rating Scale (CRS) scores. Variation in parent and teacher's CRS scores was for 38% explained by a common factor. This factor showed a heritability of 82%. The rater-specific factors showed heritabilities of 65% and 79% for parent and teacher ratings, respectively. Simonoff et al. reported a heritability of 89% for the common factor. The genetic component of this common factor was greater than in the univariate models (52% and 69-75% in teacher and maternal ratings, respectively). Derks et al. (in press) also showed a higher heritability of the common factor (78%) than of the rater specific factors (76% and 39% for maternal and teacher ratings, respectively). In summary, all three studies report a higher heritability of the common factor than of the rater-specific factors. This can be explained by the fact that when multiple indicators for a latent phenotype are used (e.g., over time or across raters), only a proportion of the measurement error of the individual ratings is passed on to the latent phenotype (Simonoff et al., 1998).Therefore, future gene finding studies could increase statistical power by focusing on the highly heritable common factor because it is less subject to measurement error.

Are the subtypes of ADHD genetically heterogeneous?

ADHD is a disorder that may include symptoms of inattention, hyperactivity/impulsivity, or both. Because of this heterogeneity in symptom profiles, concerns have been raised over the validity of the DSM-IV subtypes (Todd, 2000). In this section, we address the question if the different subtypes of ADHD are genetically heterogeneous. In other words, is the variability in symptoms profiles explained by different genetic influences on the inattentive type, the hyperactive/impulsive type and the combined type? A number of papers have looked at the familiality and heritability of the DSM-IV subtypes of ADHD. These studies failed to identify significant familial (i.e., genetic or shared environmental) clustering of the subtypes and concluded that symptom variability is largely a function of non-familial causes (Faraone et al., 2000a; Faraone et al., 2000b; Smalley et al., 2001).

Todd et al. (2001) used Latent Class Analysis (LCA; McCutcheon, 1987) to examine if the clustering of symptoms can be described with more meaningful subtypes. LCA assumes the presence of a number of latent classes with a categorical rather than a continuous distribution. Estimates are provided for: i) the number of latent classes; ii) the prevalence of each class; and iii) the item endorsement probabilities conditional on latent class membership. Todd et al. (2001) applied LCA to parent reports on 2018 female adolescent twin-pairs from the state Missouri and investigated if the original DSM-IV subtypes and the derived latent classes represent independent genetic entities. The DSM-IV

combined type and inattentive type showed a lack of familial specificity (e.g., a proband with the inattentive type has a higher chance of having a cotwin with either the inattentive or the combined type, but does not have a higher chance of having a cotwin with the hyperactive/impulsive type). The hyperactive/impulsive type did show familial specificity (e.g., a proband with the hyperactive/impulsive type has a higher chance of having a cotwin with the hyperactive/impulsive type, but does not have a higher chance of having a cotwin with the inattentive or combined type). This suggests that the hyperactive/impulsive type is independent of the other two subtypes. The LCA resulted in an eight-class solution. This eight–class solution was replicated in a sample of Australian twins (Rasmussen et al., 2002) and a similar (7-class) solution was found in an independent sample from Missouri (Volk et al., 2005). In contrast to the DSM-IV subtypes, the eight latent classes appeared to represent pure genetic categories. The authors conclude that "these results are most compatible with the presence of independent, familial forms of ADHD that are approximated by latent-class analysis and are imperfectly operationalized by DSM-IV criteria".

Is liability to ADHD continuous or categorical?

Another interesting feature of LCA is that it can help clarify whether ADHD shows a categorical or a continuous distribution. If the underlying nature of the phenotype is a continuum of problems with inattention, hyperactivity/impulsivity, or both, then symptoms endorsement profiles of the observed classes will reflect differences in severity or frequency of the reporting of symptoms only (Hudziak et al., 1998). Analyzing data on 1549 female twin-pairs, Hudziak et al. (1998) showed symptom profiles that indicated the presence of three separate continua of severity of problems: inattention, hyperactivity/impulsivity, and combined type. Thus, within the domains, the symptoms are better described as existing on a continuum rather than as discrete disease entities.

Molecular genetic studies of ADHD

Molecular genetic studies address the question which genes explain the high heritability of AHDH. It is beyond the scope of this paper to provide an extensive overview of the results of molecular genetic studies. Recently, a number of review studies on the molecular genetics of ADHD have been published (Faraone et al., 2005; Bobb et al., 2005; Thapar et al., 2005; Asherson, 2004).

Faraone et al. (2005) reviewed candidate gene studies of ADHD and computed pooled odds ratio's (ORs) across studies for gene variants examined in three or more case-control or family-based studies. Seven gene variants showed a pooled OR that is significantly larger than 1: Dopamine Receptor D4 (DRD4), Dopamine Receptor D5 (DRD5), Dopamine Transporter (DAT), Dopamine -Hydroxylase (DBH), Synaptosomal-Associated Protein 25 (SNAP-25), Serotonin Transporter (5-HTT), and Serotonin receptor (HTR1B). These small ORs are consistent with the idea that the genetic vulnerability to ADHD is mediated by many genes of small effect.

Five groups have conducted genome-wide linkage scans in an attempt to find regions of chromosomes that are involved with ADHD. We will discuss the regions for which LOD scores higher than 2 ($p<\sim.002$) were found. The first genome wide scan on ADHD was published in 2002 by Fisher et al. (2002) who analyzed data from 126 affected sibling pairs in 104 families. In 2003, the sample was extended and contained 204 families with 207 affected sibling pairs (Ogdie et al., 2003). In the extended sample, LOD > 2 was found at chromosome 16p13 and 17p11. Bakker et al. (2003) performed a genome scan on 238 children from 164 Dutch affected sib pairs with ADHD. They report a LOD score of 3.04 at chromosome 7p, and of 3.54 at chromosome 15q. Arcos-Burgos et al. (2004) analyzed data from 16 genetically isolated families in Columbia. They reported linkage peaks (LOD score > 2) at chromosomes 4q, 8q, and 11q in specific families. The fourth genome wide scan was performed in a sample of 102 families encompassing a total of 229 affected children

(Hebebrand et al., 2006). For clinical diagnosis of ADHD, the highest LOD score of 2.74 was reported on chromosome 5p. A LOD score >2 was also found at chromosome 12q. For quantitative DSM-IV measures, the highest LOD scores were observed on chromosome 5p (total and inattentive scores), and chromosome 12q (inattentive scores). For hyperactivity, no LOD scores > 2 were reported. Finally, Gayan et al. (2005) reported linkage for ADHD at chromosomes 14q32 and 20q11.

The linkage peaks of these four studies do not show much overlap. An interesting resemblance between the studies is that four genome wide scans report modest evidence (LOD > 1) for linkage at chromosome 5p. An obvious candidate gene at chromosome 5p, is the DAT gene, but in the study of Hebebrand et al., allelic variation at the DAT1 was not responsible for the linkage signal. Furthermore, the gene with the largest pooled OR as reported by Faraone et al., DRD4, is located at chromosome 11p. None of the genome wide scans reported a linkage peak at this location.

The results of these four studies are inconsistent. This may be the results of the different sampling procedures which are applied to select subjects or to differences in the definition of the phenotype. Furthermore, because each gene is expected to show a small effect, and because a correction to the type-I error (α) has to be made because of multiple testing, the statistical power in each study is low.

REFERENCE LIST

Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.

American Psychiatric Association (1994). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorder: DSM-IV (4th ed.)*. Washington, DC: American Psychiatric Association.

Arcos-Burgos, M., Castellanos, F. X., Pineda, D., Lopera, F., Palacio, J. D., Palacio, L. G. et al. (2004). Attention-deficit/hyperactivity disorder in a population isolate: Linkage to loci at 4q13.2, 5q33.3, 11q22, and 17p11. *American Journal of Human Genetics, 75,* 998-1014.

Asherson, P. (2004). Attention-Deficit Hyperactivity Disorder in the post-genomic era. *European Child and Adolescent Psychiatry, 13 Suppl 1,* I50-I70.

Bakker, S. C., Meulen van der, E. M., Buitelaar, J. K., Sandkuijl, L. A., Pauls, D. L., Monsuur, A. J. et al. (2003). A whole-genome scan in 164 Dutch sib-pairs with Attention-Deficit Hyperactivity Disorder: suggestive evidence for linkage on chromosomes 7p and 15q. *American Journal of Human Genetics, 72,* 1251-1260.

Barkley, R. A. (2002). Major life activity and health outcomes associated with attention-deficit/hyperactivity disorder. *Journal of Clinical Psychiatry, 63,* 10-15.

Baumgaertel, A., Wolraich, M. L., & Dietrich, M. (1995). Comparison of Diagnostic-Criteria for Attention-Deficit Disorders in A German Elementary-School Sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 34,* 629-638.

Beijsterveldt, C. E. M., Verhulst, F. C., Molenaar, P. C. M., & Boomsma, D. I. (2004). The genetic basis of problem behavior in 5-year-old Dutch twin pairs. *Behavior Genetics, 34,* 229-242.

Bobb, A. J., Castellanos, F. X., Addington, A. M., & Rapoport, J. L. (2005). Molecular genetic studies of ADHD: 1991 to 2004. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 132,* 109-125.

Boomsma, D. I. (2005). Sibling Interaction Effects. In B.Everitt & D. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (pp. 1831-1832). Chisester: John Wiley & Sons.

Breton, J. J., Bergeron, L., Valla, J. P., Berthiaume, C., Gaudet, N., Lambert, J. et al. (1999). Quebec Child Mental Health Survey: Prevalence of DSM-III-R mental health disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40,* 375-384.

Burt, S. A., Krueger, R. F., McGue, M., & Iacono, W. G. (2001). Sources of covariation among attention-deficit/hyperactivity disorder, oppositional defiant disorder, and conduct disorder: The importance of shared environment. *Journal of Abnormal Psychology, 110,* 516-525.

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics, 16,* 319-340.

Castellanos, F. X. & Tannock, R. (2002). Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Reviews Neuroscience, 3,* 617-628.

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the child behavior checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62,* 1017-1025.

Coolidge, F. L., Thede, L. L., & Young, S. E. (2000). Heritability and the comorbidity of attention deficit hyperactivity disorder with behavioral disorders and executive function deficits: A preliminary investigation. *Developmental Neuropsychology, 17,* 273-287.

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60,* 837-844.

Cuffe, S. P., McKeown, R. E., Jackson, K. L., Addy, C. L., Abramson, R., & Garrison, C. Z. (2001). Prevalence of attention-deficit/hyperactivity disorder in a community sample of older adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 1037-1044.

Cuffe, S. P., Moore, C. G., & McKeown, R. E. (2005). Prevalence and correlates of ADHD symptoms in the National Health Interview Survey. *Journal of Attention Disorders, 9,* 392-401.

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

Derks E.M. Dolan C.V., Hudziak J.J., Neale M.C., and Boomsma D.I. (in revision). Teacher Reports and the role of Sex and Genetics in Attention Deficit Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder (ODD). *Journal of Abnormal Psychology.*

Derks, E. M., Hudziak, J. J., Beijsterveldt van, C. E. M., Dolan, C. V., & Boomsma, D. I. (in press). Genetic Analyses of Maternal and Teacher Ratings on Attention Problems in 7-year-old Dutch Twins. *Behavior Genetics*.

Derks, E. M., Hudziak, J. J., Dolan, C. V., Ferdinand, R. F., & Boomsma, D. I. (2006). The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. *Comprehensive Psychiatry, 47,* 116-122.

Derks, E. M., Hudziak, J. J., Beijsterveldt van, C. E. M., Dolan, C. V., & Boomsma, D. I. (2004). A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. *Behavior Genetics, 34,* 571-583.

Dick, D. M., Viken, R. J., Kaprio, J., Pulkkinen, L., & Rose, R. J. (2005). Understanding the covariation among childhood externalizing symptoms: Genetic and environmental influences on conduct disorder, attention deficit hyperactivity disorder, and oppositional defiant disorder symptoms. *Journal of Abnormal Child Psychology, 33,* 219-229.

Eaves, L. J. (1976). A model for sibling effects in man. *Heredity, 36,* 205-214.

Eaves, L. J., Silberg, J. L., Meyer, J. M., Maes, H. H., Simonoff, E., Pickles, A. et al. (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 Psychology and Psychiatry and Allied Disciplines, 38,* 965-980.

Edelbrock, C., Rende, R., Plomin, R., & Thompson, L. A. (1995). A twin study of competence and problem behavior in childhood and early adolescence. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 36,* 775-785.

Eiraldi, R. B., Power, T. J., Karustis, J. L., & Goldstein, S. G. (2000). Assessing ADHD and comorbid disorders in children: the Child Behavior Checklist and the Devereux Scales of Mental Disorders. *Journal of Clinical Child Psychology, 29,* 3-16.

Faraone, S. V. & Biederman, J. (2005). What is the prevalence of adult ADHD? Results of a population screen of 966 adults. *Journal of Attention Disorders, 9,* 384-391.

Faraone, S. V., Biederman, J., & Friedman, D. (2000a). Validity of DSM-IV subtypes of attention-deficit/hyperactivity disorder: a family study perspective. *Journal of the American Acadamy of Child and Adolescent Psychiatry, 39,* 300-307.

Faraone, S. V., Biederman, J., Mick, E., Williamson, S., Wilens, T., Spencer, T. et al. (2000b). Family study of girls with attention deficit hyperactivity disorder. *American Journal of Psychiatry, 157,* 1077-1083.

Faraone, S. V., Perlis, R. H., Doyle, A. E., Smoller, J. W., Goralnick, J. J., Holmgren, M. A. et al. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry, 57,* 1313-1323.

Fisher, S. E., Francks, C., McCracken, J., McGough, J. J., Marlow, A. J., MacPhie I.L. et al. (2002). A genomewide scan for loci involved in attenion-deficit/hyperactivity disorder. *American Journal of Human Genetics, 70,* 1183-1196.

Ford, T., Goodman, R., & Meltzer, H. (2003). The British child and adolescent mental health survey 1999: The prevalence of DSM-IV disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 42,* 1203-1211.

Gaub, M. & Carlson, C. L. (1997). Gender differences in ADHD: a meta-analysis and critical review. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 1036-1045.

Gayan, J., Willcutt, E. G., Fisher, S. E., Francks, C., Cardon, L. R., Olson, R. K. et al. (2005). Bivariate linkage scan for reading disability and attention-deficit/hyperactivity disorder localizes pleiotropic loci. *Journal of Child Psychology and Psychiatry and Allied Discipline, 46,* 1045-1056.

Gillis, J. J., Gilger, J. W., Pennington, B. F., & DeFries, J. C. (1992). Attention deficit disorder in reading-disabled twins: evidence for a genetic etiology. *Jorunal of Abnormal Child Psychology, 20,* 303-315.

Gjone, H., Stevenson, J., & Sundet, J. M. (1996a). Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 588-596.

Gjone, H., Stevenson, J., Sundet, J. M., & Eilertsen, D. E. (1996b). Changes in heritability across increasing levels of behavior problems in young twins. *Behavior Genetics, 26,* 419-426.

Goodman, R. & Stevenson, J. (1989). A Twin Study of Hyperactivity .2. the Etiological Role of Genes, Family Relationships and Perinatal Adversity. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 30,* 691-709.

Gould, M. S., Bird, H., & Staghezza Jaramillo, B. (1993). Correspondence between statistically derived behavior problem syndromes and child psychiatric diagnoses in a community sample. *Journal of Abnormal Child Psychology, 21,* 287-313.

Graetz, B. W., Sawyer, M. G., & Baghurst, P. (2005). Gender differences among children with DSM-IV ADHD in Australia. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 159-168.

Graetz, B. W., Sawyer, M. G., Hazell, P. L., Arney, F., & Baghurst, P. (2001). Validity of DSM-IV ADHD subtypes in a nationally representative sample of Australian children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 1410-1417.

Hebebrand, J., Dempfle, A., Saar, K., Thiele, H., Herpertz-Dahlmann, B., Linder, M. et al. (2006). A genome-wide scan for attention-deficit/hyperactivity disorder in 155 german sib-pairs. *Molecular Psychiatry, 11,* 196-205.

Hewitt, J. K., Silberg, J. L., Neale, M. C., Eaves L.J., & Erickson, M. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics, 22,* 293-317.

Hudziak, J. J. (2001). The role of phenotypes (diagnoses) in genetic studies of attention-deficit/hyperactivity disorder and related child psychopathology. *Child and Adolescent Psychiatric Clinics of North America, 10,* 279-97, viii.

Hudziak, J. J., Copeland, W., Stanger, C., & Wadsworth, M. (2004). Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 1299-1307.

Hudziak, J. J., Derks, E. M., Althoff, R. R., Rettew, D. C., & Boomsma, D. I. (2005). The genetic and environmental contributions to attention deficit hyperactivity disorder as measured by the Conners' Rating Scales - Revised. *American Journal of Psychiatry, 162,* 1614-1620.

Hudziak, J. J., Heath, A. C., Madden, P. F., Reich, W., Bucholz, K. K., Slutske, W. et al. (1998). Latent class and factor analysis of DSM-IV ADHD: A twin study of female adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 37,* 848-857.

Hudziak, J. J., Rudiger, L. P., Neale, M. C., Heath, A. C., & Todd, R. D. (2000). A twin study of inattentive, aggressive and anxious/depressed behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 469-476.

Kuntsi, J., Rijsdijk, F., Ronald, A., Asherson, P., & Plomin, R. (2005). Genetic influences on the stability of attention-deficit/hyperactivity disorder symptoms from early to middle childhood. *Biological Psychiatry, 57,* 647-654.

Kuntsi, J. & Stevenson, J. (2001). Psychological mechanisms in hyperactivity: II The role of genetic factors. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 211-219.

Larsson, J., Larsson, H., & Lichtenstein, P. (2004). Genetic and environmental contributions to stability and change of ADHD symptoms between 8 and 13 years of age: A longitudinal twin study. *Journal of the American Academy of Child and Adolescent Psychiatry, 43,* 1267-1275.

Lavigne, J. V., Gibbons, R. D., Kaufer Christoffel, K., Arend, R., Rosenbaum, D., Binns, H. et al. (1996). Prevalence rates and correlates of psychiatric disorders among preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 204-214.

Lengua, L. J., Sadowski, C. A., Friedrich, W. N., & Fisher, J. (2001). Rationally and empirically derived dimensions of children's symptomatology: expert ratings and confirmatory factor analyses of the CBCL. *Journal of Consulting and Clinical Psychology, 69,* 683-698.

Levy, F., Hay, D. A., Bennett, K. S., & McStephen, M. (2005). Gender differences in ADHD subtype comorbidity. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 368-376.

Levy, F., Hay, D. A., McStephen, M., Wood, C., & Waldman, I. (1997). Attention-deficit hyperactivity disorder: A category or a continuum? Genetic analysis of a large-scale twin study. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 737-744.

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

Martin, N., Scourfield, J., & McGuffin, P. (2002). Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *British Journal of Psychiatry, 180,* 260-265.

McCutcheon, A. L. (1987). *Latent Class Analysis*. Newbury Park: Sage Publications.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13,* 127-143.

Meredith, W. (1993). Measurement Invariance, Factor-Analysis and Factorial Invariance. *Psychometrika, 58,* 525-543.

Muthén, L. K. & Muthén, B. O. (2000). *Mplus. Statistical analyses with latent variables. User's guide (Version 2)*. Los Angeles: Muthén & Muthén.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA: Department of Psychiatry.

Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publisher.

Neuman, R. J., Sitdhiraksa, N., Reich, W., Ji, T. H. C., Joyner, C. A., Sun, L. W. et al. (2005). Estimation of prevalence of DSM-IV and latent class-defined ADHD subtypes in a population-based sample of child and adolescent twins. *Twin Research and Human Genetics, 8,* 392-401.

Nolan, E. E., Gadow, K. D., & Sprafkin, J. (2001). Teacher reports of DSM-IV ADHD, ODD, and CD symptoms in schoolchildren. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 241-249.

Ogdie, M. N., MacPhie, I. L., Minassian, S. L., Yang, M., Fisher, S. E., Francks, C. et al. (2003). A genomewide scan for attention-deficit/hyperactivity disorder in an extended sample: Suggestive linkage on 17p11. *American Journal of Human Genetics, 72,* 1268-1279.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral Genetics*. New York.

Price, T. S., Simonoff, E., & Waldman, I. (2001). Hyperactivity in preschool children is highly heritable. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 1362-1364.

Rasmussen, E. R., Neuman, R. J., Heath, A. C., Levy, F., Hay, D. A., & Todd, R. D. (2002). Replication of the latent class structure of Attention-Deficit Hyperactivity Disorder (ADHD) subtypes in a sample of Australian twins. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 43,* 1018-1028.

Rhee, S. H., Waldman, I. D., Hay, D. A., & Levy, F. (1999). Sex differences in genetic and environmental influences on DSM-III-R attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology, 108,* 24-41.

Rietveld, M. J., Posthuma, D., Dolan, C. V., & Boomsma, D. L. (2003a). ADHD: sibling interaction or dominance: an evaluation of statistical power. *Behavior Genetics, 33,* 247-255.

Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2003b). Heritability of attention problems in children: I. Cross-sectional results from a study of twins, age 3-12 years. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 117B,* 102-113.

Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2004). Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 577-588.

Rohde, L. A., Biederman, J., Busnello, E. A., Zimmermann, H., Schmitz, M., Martins, S. et al. (1999). ADHD in a school sample of Brazilian adolescents: A study of prevalence, comorbid conditions, and impairments. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 716-722.

Saudino, K. J., Ronald, A., & Plomin, R. (2005). The etiology of Behavior problems in 7-year-old twins: substantial genetic influence and negligible shared environmental influence for parent ratings and ratings by same and different teachers. *Journal of Abnormal Child Psychology, 33,* 113-130.

Schmitz, S. & Mrazek, D. A. (2001). Genetic and environmental influences on the associations between attention problems and other problem behaviors. *Twin Research, 4,* 453-458.

Sherman, D. K., Iacono, W. G., & McGue, M. K. (1997a). Attention-deficit hyperactivity disorder dimensions: A twin study of inattention and impulsivity-hyperactivity. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 745-753.

Sherman, D. K., McGue, M. K., & Iacono, W. G. (1997b). Twin concordance for attention deficit hyperactivity disorder: A comparison of teachers' and mothers' reports. *American Journal of Psychiatry, 154,* 532-535.

Simonoff, E., Pickles, A., Hervas, A., Silberg, J. L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine, 28,* 825-837.

Smalley, S. L., McCracken, J., & McGough, J. (2001). Refining the ADHD phenotype using affected sibling pair families. *American Journal of Medical Genetics, 105,* 31-33.

Sprafkin, J., Gadow, K. D., Salisbury, H., Schneider, J., & Loney, J. (2002). Further evidence of reliability and validity of the Child Symptom Inventory-4: Parent Checklist in clinically referred boys. *Journal of Clinical Child and Adolescent Psychology, 31,* 513-524.

Stevenson, J. (1992). Evidence for a genetic etiology in hyperactivity in children. *Behavior Genetics, 22,* 337-344.

Szatmari, P., Offord, D. R., & Boyle, M. H. (1989). Ontario Child Health Study - Prevalence of Attention Deficit Disorder with Hyperactivity. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 30,* 219-230.

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,* 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,* 537-544.

Thapar, A., O'Donovan, M., & Owen, M. J. (2005). The genetics of attention deficit hyperactivity disorder. *Human Molecular Genetics, 14 Spec No. 2,* R275-R282.

Todd, R. D. (2000). Genetics of attention deficit/hyperactivity disorder: Are we ready for molecular genetic studies? *American Journal of Medical Genetics, 96,* 241-243.

Todd, R. D., Constantino, J. N., & Neuman, R. J. (2005). Redefining early-onset disorders for genetic studies: attention -deficit/hyperactiviy disorder and autism. In C.F.Zorumski & E. H. Rubin (Eds.), *Psychopathology in the genome and neuroscience era* (1 ed., pp. 33-49). Arlington: American Psychiatric Publishing, Inc.

Todd, R. D., Rasmussen, E. R., Neuman, R. J., Reich, W., Hudziak, J. J., Bucholz, K. K. et al. (2001). Familiality and heritability of subtypes of attention deficit hyperactivity disorder in a population sample of adolescent female twins. *American Journal of Psychiatry, 158,* 1891-1898.

Towers, H., Spotts, E., Neiderhiser, J. M., Hetherington, E. M., Plomin, R., & Reiss, D. (2000). Genetic and environmental influences on teacher ratings of the Child Behavior Checklist. *International Journal of Behavioral Development, 24,* 373-381.

Vierikko, E., Pulkkinen, L., Kaprio, J., & Rose, R. J. (2004). Genetic and environmental influences on the relationship between aggression and hyperactivity-impulsivity as reated by teachers and parents. *Twin Research, 7,* 261-274.

Volk, H. E., Neuman, R. J., Joyner, C., & Todd, R. D. (2005). Replication and validation of population derived ADHD phenotypes. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 138B,* 62-63.

Willcutt, E. G., Pennington, B. F., & DeFries, J. C. (2000). Etiology of inattention and hyperactivity/ impulsivity in a community sample of twins with learning difficulties. *Journal of Abnormal Child Psychology, 28,* 149-159.

Wolraich, M. L., Hannah, J. N., Baumgaertel, A., & Feurer, I. D. (1998). Examination of DSM-IV criteria for attention deficit/hyperactivity disorder in a county-wide sample. *Journal of Developmental and Behavioral Pediatrics, 19,* 162-168.

Wolraich, M. L., Hannah, J. N., Pinnock, T. Y., Baumgaertel, A., & Brown, J. (1996). Comparison of diagnostic criteria for attention-deficit hyperactivity disorder in a county-wide sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 319-324.

Young, S. E., Stallings, M. C., Corley, R. P., Krauter, K. S., & Hewitt, J. K. (2000). Genetic and environmental influences on behavioral disinhibition. *American Journal of Medical Genetics, 96,* 684-695.

CHAPTER

3

TEACHER REPORTS AND THE ROLE OF SEX AND GENETICS IN ATTENTION DEFICIT HYPERACTIVITY DISORDER (ADHD) AND OPPOSITIONAL DEFIANT DISORDER (ODD)

This chapter is in revision as: Eske M. Derks, Conor V. Dolan, James J. Hudziak, Michael C. Neale, and Dorret I. Boomsma. Teacher Reports and the role of Sex and Genetics in Attention Deficit Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder (ODD). *Journal of Abnormal Psychology, in revision.*

ABSTRACT

Attention Deficit Hyperactivity Disorder (ADHD) and oppositional defiant disorder (ODD) are more common in boys than girls. In this paper we investigate if the prevalence difference is the result of measurement bias. Secondly, we investigate sex differences in the genetic and environmental influences on variation in these behaviors. Teachers completed the Conners Teacher Rating Scale-Revised:Short version (CTRS-R:S) in a sample of 800 male and 851 female 7-year-old Dutch twins. No sex differences in the factor structure of the CTRS-R:S were found. This implies the absence of measurement bias. The heritability (56-71%) was similar in boys and girls. The specific genes that play a role are partly different in boys and girls.

INTRODUCTION

In both clinical and population samples, children diagnosed with attention deficit hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD) are predominantly male (Gaub & Carlson, 1997; Biederman et al., 2002; Loeber et al., 2000). For ADHD, the male:female ratio is estimated at 9:1 in clinical settings, and at 3:1 in the general population (Gaub & Carlson, 1997). Reviews on the literature on ODD suggest that boys show higher rates of ODD than girls, but the sex effect is modest (Loeber et al., 2000; Maughan et al., 2004). In a recent epidemiological study of 10,438 5-15 year old boys and girls, ODD was significantly more common in boys than girls (odds ratio=2.4) (Maughan et al., 2004). However, this effect was mediated by informant: no significant sex differences were found in parental reports (odds ratio=1.3), while teacher reports identified a marked preponderance of boys (odds ratio=3.9).

An important aspect in any investigation of the nature of the phenotypic sex differences is measurement invariance of an instrument with respect to sex. Mellenbergh (1989) defined measurement invariance (MI), or equivalently unbiasedness, to mean that the distribution of the observed test score, conditional on the latent construct that the test measures should be identical in each group (e.g., boys and girls). In more simple terms this means that the instrument is measuring the same construct in boys and girls (Mellenbergh, 1989; Meredith, 1993). If this is the case then we expect the score of a person to depend on that person's score on the latent construct, but not on that person's sex. If this is not the case, a boy and a girl with identical levels of problem behavior may obtain systematically (i.e., regardless of measurement error) different scores on the instrument. This is undesirable because obviously we wish our measurements to reflect accurate and interpretable differences between cases in different groups. Meredith (1993) discussed in detail the implications of Mellenbergh's definition of unbiasedness (MI) for the common factor model.

The aim of the present paper is to examine whether teacher ratings of ADHD and ODD in children are unbiased with respect to sex. Provided MI invariance can be shown to be tenable to reasonable approximation, we shall, as the second goal of this paper, estimate the genetic and environmental contributions to the phenotypic variance in ADHD and ODD in boys and girls. Lubke et al. (2004) discussed in detail the importance of establishing MI with respect to sex for the correct interpretation of any sex differences in the results of genetic modeling.

Analyses aimed explicitly at establishing MI with respect to sex, according to the manner outlined by Meredith (1993), have yet to be conducted with respect to ADHD and ODD. However, the factor structure in teacher ratings of ADHD across boys and girls has been investigated before. Fantuzzo et al. (2001) investigated the factor structure of the 28-item version of the Conners Teacher Rating Scale (CTRS) with exploratory factor analyses. They report a three-factor solution, which explained 58% of the variance. The factors admitted interpretations in terms of Conduct, Hyperactivity, and Passivity. The invariance of the factor structure with respect to sex was investigated by comparing results from random subgroups with the results from subgroups based on sex. The results supported the similarity of the factor structure across sex. A concern in this study is that the subjects in this study were 580 children from 33 classrooms located in low-income neighborhoods, and it is unclear how representative they are of the general population. In addition, there are two sources of dependence in the data: each teacher rated more than one child, and children were clustered into classes. This violates the assumption of independent observations, which is important in statistical inference.

In the present study, confirmatory factor analyses are employed on data from a large general population sample of 7-year-old twins, rated by their teacher. Two questions will be addressed. First, is the measurement model that relates differences in the latent construct of

ADHD and ODD to the observed behavior problem scores identical in boys and girls, i.e., is MI tenable? Second, do the magnitudes of the genetic and environmental influences differ, or do different genes play a role in boys and girls?

METHODS

Subjects and procedure

This study was part of an ongoing twin study of development and psychopathology in the Netherlands. The subjects were all registered at the Netherlands Twin Registry (NTR; Boomsma et al., 2002). As part of this study, we assessed a sample of Dutch twins from the birth cohorts 1992-1996, whose teachers reported on their behavior when they were seven years old.

The twins at age three are representative of Dutch three-year-old children with respect to their scores on measures such as the Child Behavior Checklist (CBCL; Oord van den et al., 1995). The Socioeconomic Status of the parents of the twins was somewhat higher than the level in the general Dutch population (Rietveld et al., 2004). When twins reached age 7 years, parents were asked to provide informed consent to approach the teacher. Consent was given by 80.1% of the parents. Teachers of these pairs received a questionnaire by mail, and were asked to return it to the NTR by mail. The response rate of the teachers was 80.0%. CTRS data were available for at least one twin in 1651 twin-pairs (1511 complete pairs).

The maternal CBCL-AP scores at age 7 years were not significantly different between families in which parents provided permission to approach the teachers (mean=2.95, sd=2.93) and families in which parents did not provide permission (mean=3.15, sd=3.18) ($t(3063)=2.0$, $p=.133$). However, mean maternal AP ratings were significantly higher for twins whose teachers did not return the questionnaire (mean=3.34; sd=3.13) than for twins whose teachers returned the questionnaire (mean=2.78; sd=2.81) ($F(1)=16.82$, $p<.001$).

To avoid biased test results due to statistical dependency of the twins, we randomly included the score from only one of the members of a twin-pair in the confirmatory factor analyses. The resulting sample consists of 1651 individual twins (800 boys and 851 girls). In the genetic analyses, the statistical dependency of the twin data is part of the genetic model and we therefore included all complete twin-pairs. Data were available for both members of a twin-pair in 1511 monozygotic (MZ) and dizygotic (DZ) twin-pairs. The number of pairs by zygosity is 248 MZ male, 251 DZ male, 294 MZ female, 234 DZ female, and 484 DZ opposite sex pairs. Some twins were rated by the same teacher (877 pairs;58%) while the remaining twins were rated by different teachers (634 pairs, 42%).

Measure

The Conners' Teacher Rating Scale-Revised (CTRS-R) is a widely used instrument to assess behavior problems (Conners, 2001; Conners et al., 1998b). The CTRS-R was developed by factor analyzing a large set of items, and including items that load highly on interpretable common factors. In addition to the scales that were derived based on factor analysis, an ADHD-index was included. This index was not developed by factor analysis, but comprises the best twelve items for distinguishing children with ADHD from children without ADHD as assessed by the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 1994; Conners, 2001). The short version, which was used in the present study, contains 28 items from the following scales: Oppositional (ODD, 5 items); Cognitive problems-Inattention (IN, 5 items); Hyperactivity (HI, 7 items); and the ADHD-index (ADHD-I, 12 items). One of the items (item 27; Excitable, impulsive) is included in both the Hyperactivity scale and the ADHD-index. The items are rated on a 4-point Likert scale for symptom severity (i.e., 0 = 'Not true at all'; 1 = 'Just a little true'; 2 = 'Pretty much

true'; 3 = 'Very much true'). The internal consistency and stability of the CTRS-R:Short version (CTRS-R:S) are good, as the Cronbach's alpha coefficients range from .88 to .95, and 6-8 week test-retest correlations range from .72 to .92 (Conners, 2001).

Distribution of the items

Because of the categorical nature of the item scores, Pearson product moment correlations underestimate the correlation of the underlying latent trait, and consequently the parameter estimates obtained in factor analyses or principal component analyses based on the Pearson correlation matrices are biased (Dolan, 1994). We therefore, adopted an approach that is suitable for factor analyzing discrete item scores. Polychoric correlations between items were obtained based on the liability threshold model (Lynch & Walsh, 1998). In the case of a 4-point Likert scale, three thresholds divide the latent liability distribution into four categories.

Criteria of measurement invariance

The criteria of MI are empirically testable in the common factor model (Meredith, 1993). Factor analysis may be viewed as a regression model in which observed variables (e.g., item scores) are regressed on a latent variable or common factor. In terms of this regression, the MI criteria are: 1) equality of regression coefficients (i.e., factor loadings) over groups; 2) equality of item intercepts over groups (i.e., differences in item means can only be the result of differences in factor means), and 3) equality of residual variances (i.e., variance in the observed variables, not explained by the common factor) over groups. When satisfied, these restrictions ensure that any differences in the mean and variance of the observed variables are due to differences in the mean and variance of the common factor.

In case of ordinal data, MI can be tested by constraining the thresholds, factor loadings, and residual variances to be equal in boys and girls. These constraints allow us to estimate group differences in the means and variances of the common factor. To this end, the mean and variance of the common factor are fixed at 0 and 1, respectively, in an arbitrary reference group. We chose to estimate the mean and variance in girls, and to use boys as the reference group (the results are identical when girls are used as the reference group). In doing so, we are modeling the observed group differences as a function of the differences in the latent liability.

Statistical analyses

The polychoric correlation matrices of the items of the four subscales were calculated using the computer program Prelis (Jöreskog & Sörbom, 1996b). All confirmatory factor analyses were performed on raw data using the program Mx (Neale et al., 2003). In principle, factor analysis of p ordinal items can be carried out using Full Information Maximum Likelihood (FIML) estimation. However, this requires repeated integration of the p-variate normal distribution, which can become computationally demanding even with as few as 12 items. We therefore chose to estimate the model parameters using Marginal Maximum Likelihood estimation (MML; Bock & Aitkin, 1981). MML maximizes the likelihood of the data conditional on the latent trait, in contrast to FIML, which maximizes the unconditional likelihood. The advantage of MML compared to FIML is that it is computationally much less demanding.

To test if the Conners' scales are MI with respect to sex, the factor structure was constrained to be identical in boys and girls. While ODD, IN and HI resulted from factor analyses, a single factor was fit to these scales. In contrast, the ADHD-I contains items related to problems with inattention and hyperactivity, and thus a two-factor model was fit. In order to detect prevalence differences in ADHD and ODD across sex, the means and variances of the latent factors were constrained to be equal in boys and girls.

The fit of the models was compared by chi square (χ^2) tests, with an (type I error probability of the statistical test) of .01. A disadvantage of the χ^2 test was noted by Browne et al. (2002) who showed that χ^2 is influenced by the unique variances of the items. If a trait is measured reliably, the inter-correlations of the items are high, and the unique variances are small. In this case, the χ^2 test may suggest a poor fit even when the residuals between the expected and observed covariances are trivial. The Standardized Root Mean-square Residual (SRMR; Bentler, 1995) is a fit index that is not sensitive to the size of the correlations. To avoid the rejection of a simpler model due to high inter-item correlations, we only reject a model if a significant χ^2 test is accompanied by large residuals (SRMR>.08; Hu & Bentler, 1999).

In addition to investigating MI with respect to sex, we will investigate sex differences in the genetic and environmental influences on the individual differences in the sum scores of the scales, given that MI is tenable (Lubke et al., 2004). The polychoric twin-correlation matrices of the four scales were calculated by zygosity and sex using the computer program PRELIS (Jöreskog & Sörbom, 1996). However, the actual genetic analyses were performed in Mx (Neale et al., 2003).

Using data from MZ and DZ twins, the variance in behavior can be attributed to genetic and environmental factors. In our sample, 58% of the twins were in the same classroom and 42% of the twins were in different classrooms. Correlations may be higher when children are rated by the same teacher. Therefore a correlated errors model developed by Simonoff et al. (1998) was used to analyze the data. In this model, individual differences in behavior are explained by additive genetic factors (A), common environmental factors that are shared between two twins of a pair (e.g., home environment; C) and unique environmental factors (E) (i.e., factors that may cause differences between twins). The unique environmental factors are allowed to correlate in twins who are placed into the same classroom, and do not correlate in twins who are placed into different classrooms. With this approach, an unbiased estimate of the genetic and environmental influences is obtained. For the genetic analyses, the items of each subscale were summed, and the sum-score was recoded in such a way that three thresholds divide the latent liability distribution into four categories. The thresholds were chosen in such a way that the categories contain more or less equal numbers of subjects. We preferred this procedure to the analysis of the raw sum scores, because the latter are skewed, and therefore cannot be subjected to maximum likelihood based on the assumption of normality (e.g., see Derks et al., 2004). The discrete data can be submitted to maximum likelihood estimation based on the discrete factor model, where as above we assume that the latent liability is normally distributed.

Sex differences are examined in two ways. First, we investigated if the estimates of the genetic and environmental variances are equal in boys and girls. This was examined by comparing the fit of a model in which the genetic and environmental variances are allowed to be different in boys and girls with the fit of a model in which they are constrained to be equal. Second, we investigated if the same genes influence phenotypic variation in boys and girls. These qualitative sex differences were evaluated by constraining the genetic correlation in opposite sex twins at .5 (similar as in same-sex DZ twins). If different genes play a role in boys and girls, the genetic correlation is expected to be lower than .5 in opposite sex twins.

RESULTS

Polychoric correlations of the subscales

The item scores within each of the four scales are highly correlated. The polychoric correlations between the items of each subscale are reported in Tables 1a-d. As expected, the items of the ADHD-index formed two subgroups. Items 1, 14, 16, 19, 25, and 26 all refer to inattention, and were included within the first subgroup. The second group of items, items 5, 9, 12, 23, 27, and 28, loaded on a hyperactive factor. The correlations were consistent with the hypothesized two-factor structure, because the correlations between items within the same factor were higher than the correlations between items of different factors.

Table 1a. Polychoric correlations Oppositional

Items	2.	6.	10.	15.	20.
2. Defiant	1.00				
6. Defies	.85	1.00			
10. Spiteful	.68	.72	1.00		
15. Argues	.81	.83	.66	1.00	
20. Explosive	.80	.68	.70	.66	1.00

Table 1b. Polychoric correlations Cognitive problems-Inattention

Items	4.	8.	13.	18.	22.
4. Forgets things	1.00				
8. Poor spelling	.74	1.00			
13. Poor reading	.66	.77	1.00		
18. Lacks interest	.63	.56	.49	1.00	
22. Poor arithmetic	.75	.66	.57	.53	1.00

Table 1c. Polychoric correlations Hyperactivity

Items	3.	7.	11.	17.	21.	24.	27.
3. Restless	1.00						
7. Always on the go	.71	1.00					
11. Leaves seat	.72	.71	1.00				
17. Difficulty waiting	.64	.72	.76	1.00			
21. Runs about	.67	.73	.77	.67	1.00		
24. Difficulty playing	.67	.72	.74	.78	.79	1.00	
27. Excitable	.65	.74	.73	.78	.75	.80	1.00

Table 1d. Polychoric correlations ADHD-index

	Items	1.	14.	16.	19.	25.	26.	5.	9.	12.	23.	27.	28.
AP	1. Inattentive	1.00											
	14. Attention span	.90	1.00										
	16. Only pays attention	.59	.62	1.00									
	19. Distractibility	.93	.94	.61	1.00								
	25. Fails to finish	.76	.79	.63	.79	1.00							
	26. Not follow instructions	.72	.76	.62	.73	.83	1.00						
HI	5. Disturbs	.69	.61	.51	.64	.60	.57	1.00					
	9. Remain still	.70	.70	.56	.71	.64	.60	.79	1.00				

Note:AP=Attention Problems; HI=Hyperactivity

Table 1d. Polychoric correlations ADHD-index *(Continued)*

ITEMS	1.	14.	16.	19.	25.	26.	5.	9.	12.	23.	27.	28.
12. Fidgets	.65	.67	.47	.66	.57	.53	.63	.81	1.00			
23. Interrupts	.60	.55	.53	.60	.54	.50	.85	.76	.62	1.00		
27. Excitable	.61	.58	.53	.59	.52	.52	.73	.79	.63	.76	1.00	
28. Restless	.67	.65	.51	.68	.58	.53	.76	.90	.82	.75	.83	1.00

Note:AP=Attention Problems; HI=Hyperactivity

MEASUREMENT INVARIANCE

We tested for MI by constraining the factor loadings, thresholds, and residual covariance matrices to be equal for boys and girls while allowing the factor means and variances to be different. The factor structure of ODD was MI with respect to sex ($\chi^2(18)=16.66$, $p>.10$; SRMR=.01 and SRMR=.06 in boys and girls respectively). MI was also tenable for the ADHD-I ($\chi^2(55)=70.41$, $p>.05$; SRMR=.03 and SRMR=.05 in boys and girls respectively). IN and HI both showed statistically significant different factor structures in boys and girls ($\chi^2(18)= 98.45$, $p<.001$, and $\chi^2(26)= 57.99$, $p<.001$, respectively). However, the residuals between expected correlation matrices under the constrained and the non-constrained model were small (SRMR=.01 in girls, and SRMR=.02 in boys, for both IN and HI). Apparently, the lack of fit was the result of the high inter-correlations between the items, and not of high residuals between the expected covariance matrices. Therefore, we did not reject the model in which MI is assumed to be tenable. Table 2 provides the factor loadings and thresholds of the best-fitting model.

Table 2. Promax rotated factor loadings and thresholds (T) of the best-fitting factor model

	ITEMS	LOADINGS FACTOR 1	LOADINGS FACTOR 2	T 1	T 2	T 3
Oppositional	2. Defiant	.91	-	.71	1.65	2.45
	6. Defies	.88	-	1.02	2.00	2.74
	10. Spiteful	.69	-	1.49	2.32	3.13
	15. Argues	.84	-	.82	1.80	2.52
	20. Explosive	.78	-	1.20	1.92	3.12
Cognitive problems-Inattention	4. Forgets things	.93	-	.34	1.29	1.82
	8. Poor spelling	.81	-	.27	.95	1.44
	13. Poor reading	.71	-	.39	.87	1.21
	18. Lacks interest	.70	-	1.01	1.65	2.27
	22. Poor arithmetic	.82	-	.55	1.22	1.68
Hyperactivity	3. Restless	.73	-	.10	.94	1.51
	7. Always on the go	.79	-	.66	1.31	1.87
	11. Leaves seat	.82	-	.62	1.37	1.83
	17. Difficulty waiting	.83	-	.12	.91	1.58
	21. Runs about	.79	-	1.21	1.89	2.51
	24. Difficulty playing	.87	-	.34	1.17	2.03
	27. Excitable	.86	-	.40	1.12	1.84

Note: AP=Attention Problems; HI=Hyperactivity

Table 2. Promax rotated factor loadings and thresholds (T) of the best-fitting factor model *(Continued)*

		Items	Loadings factor 1	Loadings factor 2	T 1	T 2	T 3
ADHD-index	AP	1. Inattentive	.99	-.09	-.36	.76	1.47
		14. Attention span	.95	-.03	.03	.87	1.55
		16. Only pays attention	.45	.23	.36	1.20	1.88
		19. Distractibility	.94	-.01	.59	1.54	2.25
		25. Fails to finish	.72	.09	.24	1.27	1.99
		26. Not follow instructions	.68	.09	.38	1.30	1.86
	HI	5. Disturbs	.05	.80	.20	1.03	1.52
		9. Remain still	.05	.88	.92	1.68	2.24
		12. Fidgets	.16	.67	.12	.94	1.53
		23. Interrupts	-.05	.85	.54	1.37	2.07
		27. Excitable	-.09	.91	.46	1.15	1.83
		28. Restless	-.06	.97	.50	1.19	1.81

Note: AP=Attention Problems; HI=Hyperactivity

The thresholds are constrained to be equal in boys and girls. The means and variances of the latent factors in boys are constrained at 0 and 1, respectively. The means and variances of the latent factors are freely estimated in girls. The variances in girls were not significantly different from 1. The means of the latent factors are estimated at -.60 (oppositional), -.86 (hyperactivity), -.42 (ADHD-I AP), and -.52 (ADHD-I-HI). The mean of the cognitive problems-inattention factor is not significantly different between boys and girls.

Differences in the liability distribution among boys and girls

The present results suggest that the four scales of the CTRS-R:S which are related to ADHD and ODD are MI with respect to sex. This implies that differences between boys and girls are attributable to differences in the latent factors. Therefore, we were able to test whether the prevalence of ADHD and ODD is different between boys and girls. To this end, we equated both the mean and the variance of the latent factors. The variances of the latent factors were not significantly different between boys and girls for any one of the four rating scales (ODD: $\chi^2(1)=3.26$; IN: $\chi^2(1)=.621$; HI: $\chi^2(1)=3.814$; ADHD-index: $\chi^2(2)=6.85$, all $p>.01$). In contrast, the means of the latent factors were significantly higher in boys than girls for ODD ($\chi^2(1)=45.94$, $p<.001$), HI ($\chi^2(1)=150.36$, $p<.001$), and the ADHD-index ($\chi^2(2)=165.68$, $p<.001$). The mean differences were .60 for ODD, .86 for HI, and .42 and .52 for the inattention and hyperactivity factor of the ADHD-I, respectively. Because the variances of the common factors equal 1 in both groups, these mean differences are expressed in standard deviation units. The common factor means of IN were not significantly different across sex ($\chi^2(1)=1.96$, $p>.10$).

Genetic analyses

Having established measurement invariance of the CTRS-R:S scales with respect to sex, we proceeded to investigate the genetic and environmental contributions to the phenotypic individual differences in the discretized sum-scores. The twin-correlations are shown by zygosity and sex in Table 3, for same and different teacher separately. All correlations are higher in MZ twins than in DZ twins, which suggests the presence of genetic influences. The correlations are higher for twin pairs rated by the same teacher than for twin pairs rated by different teachers. This was taken into account by allowing a correlation between the unique environmental factors in twins rated by the same teacher. The lower correlations in opposite-sex DZ twins than in same-sex DZ twins show that different genes may play a role in boys and girls.

Table 3. Polychoric twin-correlations of the Conners Teacher Rating Scales rated by Same Teachers (ST) versus Different Teachers (DT)

	Oppositional		Cognitive problems-Inattention		Hyperactivity		ADHD-index	
	ST	DT	ST	DT	ST	DT	ST	DT
Monozygotic males	.86	.53	.90	.76	.81	.59	.85	.59
Dizygotic males	.50	.49	.64	.29	.42	.14	.49	.19
Monozygotic females	.87	.43	.92	.64	.83	.47	.87	.52
Dizygotic females	.66	.10	.60	.42	.34	.25	.45	.24
Opposite sex twins	.37	.13	.44	.23	.30	.17	.37	.27

Model fitting analyses showed that variation in all four scales could be explained by additive genetic and unique environmental effects. The influences of the shared environment were not significantly different from zero. The magnitude of the influences of genes and environment did not significantly differ between boys and girls. The standardized estimates of genetic and environmental influences are shown in Table 4. Genetic effects explained 56 to 71% of the variation in the CTRS subscales. Unique environmental effects explained the remaining 29-44% of the variation. For all four scales, the genetic correlation was significantly lower than .5 in opposite-sex twins. This implies that different genes play a role in males and females. The genetic correlation in opposite-sex twins was .16 for oppositional behavior, .35 for cognitive problems-inattention, .21 for hyperactivity-impulsivity, and .32 for the ADHD-index.

Table 4. Standardized estimates of the genetic and environmental effects on problem behavior

	Additive genetic effects (A)	Nonshared evironmental effects (E)
Oppositional	56	44
Cognitive problems-Inattention	71	29
Hyperactivity	58	42
ADHD-index	61	39

DISCUSSION

The purpose of the present study was twofold. First, we investigated if teacher ratings on ADHD and ODD in boys and girls are influenced by measurement bias with respect to sex. Second, genetic and environmental influences on variation in ADHD and ODD were compared between boys and girls.

Measurement invariance

Teacher ratings on ADHD and ODD demonstrate MI with respect to sex. In other words, teachers assess these behavior problems equally well in boys and girls, and are not influenced by the child's sex when rating a child. Sex differences in observed scores on ADHD and ODD can therefore be interpreted as mean differences in the latent construct. This supports the contention that the reported sex differences in ADHD and ODD (Gaub & Carlson, 1997; Loeber et al., 2000; Maughan et al., 2004) are due to a higher liability for the disorder in boys than girls and not to measurement bias. In addition, the CTRS is an ideal tool to use in twin studies whereas the comparison of heritabilities in boys and girls is not hindered by measurement bias with respect to sex.

Differences in the liability distribution among boys and girls

Seven-year-old boys showed higher mean scores than girls on the teacher reported ODD and HI scales, and on the ADHD-index. Although sex differences on ADHD and ODD have been reported before both in parent reports (Levy et al., 2005) and teacher reports (Nolan et al., 2001), this study demonstrates that the differences in mean scores are related to differences in the latent constructs of ADHD and ODD. It was surprising that no sex differences were found for IN, because Conners (2001) reported higher IN scores in boys than girls. More research needs to be done investigating the sex differences in ADHD, its subtypes, and the co-occurrence with cognitive problems. In addition, future studies should reveal if the sex differences in the liability distribution vary as a function of age or informant (e.g., teacher vs parent reports).

Quantitative and qualitative differences in the heritability among boys and girls

More than half the variation in ADHD and ODD in boys and girls is explained by genetic influences. The remaining variation is explained by unique environmental influences. The magnitude of the influences of genes and environment is the same in boys and girls. However, variation in ADHD and ODD in boys and girls is explained by only partly overlapping sets of genes. The genetic correlation between DZ opposite sex twins was significantly lower than 0.5 , which is the theoretical value if the same genes influence behavior in boys and girls. The consequence of this lower genetic correlation is apparent in the lower phenotypic correlations in DZ opposite-sex twins than in DZ same-sex twins for oppositional behavior, cognitive problems-inattention, hyperactivity, and the ADHD-index.

Only a few studies have addressed quantitative and qualitative sex differences in heritability estimates from teacher ratings. Saudino et al. (2005) reported qualitative sex differences in heritability in 7-year-old twins rated on hyperactive behavior by the same teacher, i.e. they obtained evidence that different genes are expressed in boys and girls. They did not report any quantitative sex differences, which is in agreement with the current findings. Vierikko et al. (2004) report that for teacher ratings of hyperactivity-impulsivity in 12-year old twins, lower correlations in opposite-sex twins than in same-sex DZ twins were seen. However, variation in these data was influenced by genes and by significant shared environmental influences. It was not possible to decide if the lower opposite sex correlations were the result of sex-specific genetic influences or sex-specific shared environmental influences, although the presence of the latter was slightly more likely. The finding of different genetic influences in boys and girls in teacher ratings stands in contrast with results based on parental reports. In parent ratings, qualitative sex differences are not found for attention problems (Rietveld et al., 2004) or ODD (Hudziak et al., 2005). The different findings in parent and teacher ratings may be explained by the fact that the behavior of children depends on the context in which they are observed.

In the Netherlands Twin Registry, teacher data are collected at the ages 7, 10, and 12 years. The sample sizes at the ages 10 and 12 are currently relatively small. In the future, we plan to address the issue of qualitative sex differences in teacher ratings in a longitudinal framework. The results of such a study will reveal if the finding of sex-differences in the specific genes that play a role is also present in older children. A further interesting issue that could be investigated in a study that includes data from children with different ages, concerns the measurement invariance of ADHD with respect to age.

Limitations

The results of the present study should be interpreted with consideration of the following limitations. First, we did not replicate the factor structure of the CTRS-R:S by means of exploratory factor analyses of the 28 items. To take the ordinal nature of the data into account, we used the liability threshold model (Bock & Aitkin, 1981). Because of computational limitations, the number of included factors is limited. Therefore, we performed con-

firmatory factor analyses, in which we assumed that the items are correctly assigned to the four scales and that cross loadings are absent. Second, teacher ratings were shown to be measurement invariant with respect to sex but this finding may not generalize to parent ratings. The correlations between Conners parent and teacher ratings are small to moderate with a range of .18 to .52 (Conners, 2001). It has been shown that parents and teachers rate partly different aspects of the child's behavior (Derks et al., in press; Martin et al., 2002). Future studies should reveal if measurement invariance is also tenable in parent ratings. Third, the mean level of AP reported in this study slightly underestimates the true level of AP in the population because the level of AP according to the mothers was somewhat lower in teachers who did return the questionnaire than in teachers who did not return the questionnaire. Fourth, data were collected in 7-year-old twins only. As noted above, future studies should reveal if the current results generalize to children from different age groups.

CONCLUDING COMMENTS

We examined psychometric properties of the Conners Teacher Rating Scale-Revised: Short version in 7-year-old twins. The scales were measurement invariant (i.e., unbiased) with respect to sex. Therefore, the sex differences that are often reported are not caused by measurement bias, but are the result of genuine sex differences in ADHD and ODD. These results reveal that assesment of ADHD and ODD symptoms, through teacher reports on the CTRS-R:S, provide a solid methodological starting point for measuring sex differences in mean scores or in heritabilities. Variation in teacher ratings of children's problem behavior is mainly influenced by genetic factors. The size of the genetic influences did not depend on the child's sex, but partly different genes are expressed in boys and girls. Future studies should reveal if these findings generalize to children from different age groups.

REFERENCE LIST

American Psychiatric Association (1994). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorder: DSM-IV (4th ed.)*. Washington, DC: American Psychiatric Association.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Biederman, J., Mick, E., Faraone, S. V., Braaten, E., Doyle, A., Spencer, T. et al. (2002). Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *American Journal of Psychiatry, 159,* 36-42.

Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46,* 443-459.

Boomsma, D. I., Vink, J. M., Beijsterveld van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Browne, M. W., MacCallum, R. C., Kim, C., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods, 7,* 403-421.

Conners, C. K. (2001). *Conners' Rating Scales-Revised*. New York and Toronto: Multi-Health Systems, Inc.

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26,* 257-268.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research, 7,* 659-669.

Derks, E. M., Hudziak, J. J., Beijsterveldt van, C. E. M., Dolan, C. V., & Boomsma, D. I. (in press). Genetic Analyses of Maternal and Teacher Ratings on Attention Problems in 7-year-old Dutch Twins. *Behavior Genetics*.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47,* 309-326.

Fantuzzo, J., Grim, S., Mordell, M., McDermott, P., Miller, L., & Coolahan, K. (2001). A multivariate analysis of the revised Conners' Teacher Rating Scale with low-income, urban preschool children. *Journal of Abnormal Child Psychology, 29,* 141-152.

Gaub, M. & Carlson, C. L. (1997). Gender differences in ADHD: a meta-analysis and critical review. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 1036-1045.
Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.
Hudziak, J. J., Derks, E. M., Althoff, R. R., Copeland, W. M., & Boomsma, D. I. (2005). The genetic and environmental contributions to oppositional defiant behavior: A multi-informant twin study. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 907-914.
Jöreskog, K. G. & Sörbom, D. (1996). *PRELIS 2 User's Reference Guide*. Chicago: Scientific Software International.
Levy, F., Hay, D. A., Bennett, K. S., & McStephen, M. (2005). Gender differences in ADHD subtype comorbidity. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 368-376.
Loeber, R., Burke, J. D., Lahey, B. B., Winters, A., & Zera, M. (2000). Oppositional defiant and conduct disorder: A review of the past 10 years, part I. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 1468-1484.
Lubke, G. H., Dolan, C. V., & Neale, M. C. (2004). Implications of absence of measurement invariance for detecting sex limitation and genotype by environment interaction. *Twin Research, 7,* 292-298.
Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.
Martin, N., Scourfield, J., & McGuffin, P. (2002). Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *British Journal of Psychiatry, 180,* 260-265.
Maughan, B., Rowe, R., Messer, J., Goodman, R., & Meltzer, H. (2004). Conduct disorder and oppositional defiant disorder in a national sample: developmental epidemiology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 609-621.
Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13,* 127-143.
Meredith, W. (1993). Measurement Invariance, Factor-Analysis and Factorial Invariance. *Psychometrika, 58,* 525-543.
Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA: Department of Psychiatry.
Nolan, E. E., Gadow, K. D., & Sprafkin, J. (2001). Teacher reports of DSM-IV ADHD, ODD, and CD symptoms in schoolchildren. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 241-249.
Oord van den, E. J. C. G., Koot, H. M., Boomsma, D. I., Verhulst, F. C., & Orlebeke, J. F. (1995). A twin-singleton comparison of problem behaviour in 2-3-year-olds. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 36,* 449-458.
Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2004). Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 577-588.
Saudino, K. J., Ronald, A., & Plomin, R. (2005). The etiology of Behavior problems in 7-year-old twins: substantial genetic influence and negligible shared environmental influence for parent ratings and ratings by same and different teachers. *Journal of Abnormal Child Psychology, 33,* 113-130.
Simonoff, E., Pickles, A., Hervas, A., Silberg, J. L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine, 28,* 825-837.
Vierikko, E., Pulkkinen, L., Kaprio, J., & Rose, R. J. (2004). Genetic and environmental influences on the relationship between aggression and hyperactivity-impulsivity as reated by teachers and parents. *Twin Research, 7,* 261-274.

CHAPTER

4

THE RELATIONS BETWEEN DISC-IV DSM-DIAGNOSES OF ADHD AND MULTI-INFORMANT CBCL-AP SYNDROME SCORES

This chapter is published as: Eske M. Derks, James J. Hudziak, Conor V. Dolan, Robert F. Ferdinand, Dorret I. Boomsma (2006). The relations between DISC-IV DSM-diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. *Comprehensive Psychiatry, 47,* 116-122.

ABSTRACT

Previous studies have examined the relation between Attention Problems (AP) obtained with the Child Behavior Checklist (CBCL), and ADHD assessed with the Diagnostic and Statistical Manual of Mental Disorders (DSM). We will examine this relation across sex using multi-informant data. Parents of 12,538 twins completed questionnaires at ages 3, 5, 7, 10, and 12 and teachers at age 10. The mothers of a sample of 283 boys and 291 girls who scored either low or high on longitudinal maternal CBCL-AP, were interviewed. Children with a low AP-score obtained a negative ADHD diagnosis in 96% of cases. Children with a high AP-score obtained a positive diagnosis in 36% (girls) and 59% (boys) of cases. The association between paternal and maternal AP-ratings and ADHD was the same, while the association between teacher AP-ratings and ADHD was low. The association between AP and ADHD is higher in boys than girls, possibly due to a bias towards the male manifestation of ADHD.

INTRODUCTION

Attention Deficit Hyperactivity Disorder (ADHD) is defined by impaired attention skills, impulsivity, and hyperactivity across the lifespan (American Psychiatric Association, 1994). It is one of the most commonly diagnosed behavioral disorders in children: the prevalence rates usually lie between 3 and 5% (Bradley & Golden, 2001). ADHD is diagnosed in a variety of ways, including both checklists and interviews. Interview methods are time-consuming and costly, and are therefore difficult to use in large-scale studies. For this reason, epidemiological studies often use behavior checklists.

Several studies have shown that behavior checklists and DSM-interviews measure overlapping constructs (Biederman et al., 1993; Chen et al., 1994; Gould et al., 1993; Gould et al., 1993; Jensen & Watanabe, 1998; Kasius et al., 1997; Kazdin & Heidish, 1984; Steinhausen et al., 1997). Correlations between Attention Problem (AP)-scores and the number of DSM-symptoms are moderate to high (Achenbach & Rescorla, 2001). As expected, the correlations are higher when both measures are obtained from the same informant (the parents) than when measures are obtained from different informants (parents vs. children) (de la Osa et al., 1997). In addition, checklist scores discriminate between children with and without a DSM diagnosis of ADHD. Steinhausen, Winkler Metzke, Meier, and Kannenberg (1997) studied a population sample of 6- to 17-year-old children and compared the mean CBCL-AP scores of 272 control children, and 35 children diagnosed with ADHD. Children with ADHD scored higher on parental AP than controls.

Other measures of association are the positive and negative predictive power, sensitivity, and specificity. Positive predictive power (PPP) refers to the proportion of children with a high CBCL score who obtain a positive DSM diagnosis (i.e., diagnosed as affected), and negative predictive power (NPP) refers to the proportion of children with a low CBCL score who obtain a negative DSM diagnosis (i.e., unaffected). Sensitivity and specificity refer to the proportion of children with a positive DSM diagnosis, who score high on the CBCL, and the proportion of children with a negative DSM diagnosis, who score low on the CBCL, respectively.

Hudziak, Copeland, and Stanger (2004) examined the sensitivity and specificity of ADHD at different CBCL-AP cutpoints in 101 male and 82 female siblings of referred children. They found that a T-score of 55 minimized the number of false positives and false negatives. The PPP and NPP were both high: 80% and 90%, respectively.

A drawback of the Diagnostic Efficiency Measures (DES) is their dependence on the baseline prevalence of the disorder. This makes it difficult to compare the results of studies with different sampling procedures. Therefore, we summarized the DES statistics and the baseline prevalences of six studies that examined the convergence of CBCL-AP and ADHD in Table 1. In some studies, the DES statistics were not reported directly and the statistics were derived from the reported number of children who scored below or above the CBCL-cutpoint by diagnosis. More details about these calculations may be obtained from the first author.

Table 1. Diagnostic Efficiency Statistics of studies that examined the association between behavior checklist scores and ADHD

STUDY	SAMPLE	N BOYS/GIRLS	CUTPOINT	PREVALENCE (%)	PPP	NPP	SE	SP
Gould et al. (1993)	NR	157	T > 65	23	.36	.96	.46	.95
Chen et al. (1994)	SR	111/108	T 65	16/8	1.00 (boys) .67 (girls)	.86 boys) .93 (girls)	.17 (boys) .22 (girls)	1.00 (boys) .99 (girls)
Eiraldi et al. (2000)	R	192/50	T 65	83	.93	.37	.78	.69
Lengua et al. (2001)	R	203	Based on regression	29	.50	.71	.02	.99
Sprafkin et al. (2002)	R	247/0	T 60	71	.78	.83	.97	.33
Hudziak et al. (2004)	SR	101/82	T 65	36	.97	.76	.47	.99

R=clinically referred sample; NR=non-referred sample; SR=siblings of referred children; PPP=Positive Predictive Power, NPP=Negative Predictive Power, SE=Sensitivity, SP=Specificity

As can be seen in Table 1, Gould, Bird, and Staghezza Jaramillo (1993) found the lowest PPP. This may in part be due to a lower baseline prevalence in her sample. The chance of being diagnosed positive for ADHD was 23% in the total sample and 36% in children with a high CBCL score. The studies of Eiraldi et al. (2000) and Sprafkin et al. (2002) have been conducted on clinically referred samples, and therefore report high baseline prevalences of ADHD (83% and 71% respectively). They observed a PPP of .93 and .78, respectively. The high prevalence of ADHD decreases the chance of a false positive case, but the specificity of the DSM-diagnosis is much lower than in the Gould et al. study. These comparisons emphasize the fact that the value of a single statistic does not provide information about the association between a test and a diagnosis, while the combination of all four statistics does.

A low sensitivity was seen in the Lengua study (2001). An examination of the 2×2 contingency table revealed that the logistic regression analyses predicted a positive DSM-diagnosis in only 2 cases. The Lengua study illustrates that CBCL-scores may not improve the prediction of ADHD if the PPP is lower than 50%. In summary, the convergence between AP and ADHD is moderate to high, depending on the kind of sample that is used. Due to differences in baseline prevalences, the positive predictive power is higher in referred than in non-referred samples.

Most studies that report on the convergence between checklists and DSM diagnoses have used the Child Behavior Checklist (CBCL). However, AP-scales of the Devereux Scale of Mental Disorders, and the Child Symptom Inventory-4 have also been shown to be good predictors of ADHD (Eiraldi et al., 2000; Eiraldi et al., 2000).

Although the convergence between AP and ADHD has been the focus of previous research, several questions remain unanswered. Firstly, is the association between behavior checklists and ADHD similar in boys and girls? Chen, Faraone, Biederman, and Tsuang (1994) found a higher positive predictive power in boys than girls, but the significance of this difference was not tested formally. Secondly, most studies have used clinically referred samples. It is not clear to what extent the results of these studies generalize to population samples. The study of Gould et al. (1993) is the only study which selected children from a general population. The association in this study was lower than the association in studies that examined referred samples. Finally, none of the studies included maternal, paternal, and teacher ratings. The inclusion of different informants might improve the prediction of DSM-diagnoses, because each informant observes the child in a slightly different situation and may have unique interactions with the child.

The purpose of this study is to further investigate the association between CBCL-AP and DSM-IV ADHD. Equal numbers of boys and girls were selected from a general population sample on the basis of longitudinal maternal CBCL-scores. Half of the subjects scored low on AP (T-score below 55; controls), and the other half of the subjects scored high on AP (T-score above 65; probands). The selection of probands diminishes the problem of under representation of affected children. Data on maternal, paternal, and teacher rated AP were collected when the children were 7, 10, and 12 years old. Mothers were interviewed when the child was around the age of 12. Shortly after this diagnostic interview, mothers completed the CBCL. We addressed the following issues. Firstly, is the cross-sectional association between AP and ADHD different in boys and girls? Secondly, can the prediction of ADHD be improved if AP-scores from multiple informants are available?

METHODS

SUBJECTS

The subjects are Dutch twins whose parents voluntarily registered with the Netherlands Twin Registry when the twins were born (Boomsma et al., 2002). 12,538 twins are currently participating in a longitudinal study, in which surveys are sent to their parents and teachers. In this longitudinal study, mothers and fathers are asked to complete the CBCL. For the present study, we analyzed parental data of 10-year-old twins from cohorts 1986-1993. In addition, we included teacher data of 10-year-old twins from cohorts 1989-1991.

In the present study, 356 families from the cohorts 1989-1992 were selected based on the maternal AP-scores obtained at age 7, 10, and 12 years. These families received a letter in which the mother of the twins was asked to participate with a structured interview study. Of these 356 families, 287 agreed to participate (80.6%), 64 families refused (18.0%), and 5 families did not respond, and could not be contacted by phone (1.4%). The sample of 287 twin-pairs consisted of 283 males and 291 females. At the time of the interview, the twins were 10 to 13 years old with a mean of 11.99 years.

SELECTION

Subjects were selected on the basis of their standardized maternal CBCL ratings (T-scores) at the ages 7, 10, and 12 years. Subjects were excluded if maternal ratings were available only at one time-point, or if they suffered from a severe handicap, which disrupts daily functioning. Twin-pairs were selected if at least one of the twins scored high on AP (probands) or if both twins scored low on AP (controls). A high score was defined as a T-score above 60 at all available time-points (age 7, 10, and 12) and a T-score above 65 at least once. A low score was defined as a T-score below 55 at all available time-points. The control twins were matched with proband twins on the basis of sex, cohort, maternal age, and Social Economic Status (SES). The criteria resulted in the selection of three types of children: children who scored low (controls), children who scored high (probands) and children who obtained an intermediate score (intermediate group). T-scores were computed in boys and girls separately. In other words, girls were selected if they scored low or high compared to other girls, and boys were selected if they scored low or high compared to other boys. This procedure resulted in the selection of an equal number of boys and girls (283 and 291 respectively). The mean AP-scores before and after selection are reported in Table 2.

Table 2. Mean maternal AP-scores before and after selection

			N	Mean maternal AP score (SD)		
				Age 7	Age 10	Age 12
Boys	Before selection	Total	6191	3.40 (3.12)	3.50 (3.32)	3.09 (3.16)
	After selection	Proband group	56	10.58 (2.98)	11.20 (2.69)	12.00 (4.77)
		Intermediate group	53	6.73 (2.84)	6.71 (2.14)	4.11 (2.64)
		Control group	174	1.71 (1.41)	1.90 (1.56)	2.27 (2.20)
Girls	Before selection	Total	6347	2.49 (2.66)	2.46 (2.66)	2.06 (2.36)
	After selection	Proband group	84	8.40 (2.22)	8.35 (2.90)	8.00 (2.27)
		Intermediate group	46	4.83 (1.70)	5.11 (2.49)	3.33 (1.21)
		Control group	161	1.18 (1.11)	1.05 (.99)	1.35 (1.56)

Procedure

The selected families received a letter inviting them to participate. Mothers of twins, who agreed to participate and who returned the informed consent form by mail, were contacted by phone to schedule the interview. At the agreed date and time, the mother was interviewed by phone. Within four months after the interview, the mother completed a CBCL, which she received by mail.

Measures

The Child Behavior Checklist (Achenbach, 1991a,b; Achenbach, 1992) is a standardized questionnaire for parents to report the frequency and intensity of behavioral and emotional problems exhibited by their child in the past six months. The CBCL 4/18 was completed by both parents when the children were 10 years old, and by the mother within four months after the interview. The Attention Problem scale (11 items) was used as a predictor for ADHD (Verhulst et al., 1996).

Teachers completed the Teacher's Report Form (TRF) (Achenbach, 1991c) when the children were 10 years old. The Dutch Attention Problem scale (20 items) was used in the present study. Ten items of the CBCL-AP scale and the TRF-AP scale overlap.

The Diagnostic Interview Schedule for Children (DISC-IV) (Shaffer et al., 1993) is a structured diagnostic interview. It can be used to assess the presence of DSM-IV diagnoses, including Attention-Deficit Hyperactivity Disorder (ADHD). The Dutch translation was obtained from Ferdinand and Ende van der (1998). The mothers of twins were interviewed by two experienced research assistants to determine which symptoms of ADHD were displayed by the twins during the last year. A child was diagnosed positive for ADHD if he or she met type-A criteria of the DSM-IV. No distinction was made between ADHD of the primarily Inattentive type and ADHD of the primarily Hyperactive/Impulsive type. Two-hundred-forty-eight interviews were audiotaped. To assess the quality of the interviews, a research assistant, who was blind to the results of the interview, listened to 40 of these interviews. It was established that the interviewers had made no mistakes that could have altered the results of the interview, and the number of ADHD symptoms as scored by the interviewer and the research assistant were the same for all subjects.

Statistical analyses

All statistical analyses were performed on sum scores in SPSS/Windows 11.0 (SPSS, 2001). The total sample of 287 twin-pairs was divided into two samples to avoid dependency inherent in twin and sibling data. To this end, the two members of a twin-pair were randomly assigned to a different sample. The descriptive statistics are reported for the total sample, but the statistical tests were performed in each sample separately. This enabled a cross-validation of the results, although the cross-validation sample does not provide a completely independent replication, because of the correlation between the twins.

Positive and Negative Predictive Power, sensitivity and specificity were computed to examine the cross-sectional association between the maternal CBCL AP score that was obtained shortly after the interview and ADHD. Children who obtained an intermediate score on the CBCL were excluded from the sensitivity analyses. They were included in all remaining analyses. In order to examine the discriminative power of the CBCL, mean AP scores of children with and without ADHD were computed. A two-way ANOVA was used to test for effects of diagnosis and sex. The type-I error rate was corrected for multiple testing in two ways. Firstly, the (type I error probability) of each test was set at .01, which is more stringent than the usual of .05. Secondly, an effect was only assumed to be present if it was significant given is .01 in both random samples of twins. In addition, correlations between AP-scores and number of ADHD-symptoms were calculated. These correlations were corrected for restriction of range due to selection with Pearson-Lawley selection rules (Lawley, 1943).

Linear stepwise regression analyses were performed to examine whether the inclusion of paternal or teacher data improved the prediction of the number of ADHD symptoms. An advantage of regression analysis is the possibility of including covariates such as sex, and of investigating associated interaction effects (e.g., between sex and CBCL scores). A significant interaction term suggests that the predictive value of AP is different in boys and girls.

RESULTS

Predictive power, specificity, and sensitivity

The cross-sectional association between CBCL-AP and DSM-ADHD was examined by calculating the predictive power of AP, and its specificity and sensitivity. The number of children by diagnosis and by CBCL-high vs. CBCL-low is reported for each sex in Table 3a. Table 3b provides the Diagnostic Efficiency Statistics. The high NPP in boys and girls suggests that a low score on the CBCL almost perfectly predicts a negative diagnosis for ADHD. In contrary, a high score on the CBCL correctly predicts the presence of ADHD in 59% of the boys and 36% of the girls. Log-linear tests showed that the percentages of boys and girls in the four categories were significantly different ($p<.05$ in both samples). Boys with a high AP-score are more often diagnosed positive for ADHD than girls with a high AP-score.

Table 3a. The number of children who obtain a diagnosis for ADHD by CBCL high (T > 65) and CBCL low (T < 55)

	Boys			Girls		
	No ADHD	ADHD	Total sample	No ADHD	ADHD	Total sample
CBCL low	151	7	158	155	5	160
CBCL high	14	20	34	36	20	56
Total sample	165	27	192	191	25	216

Table 3b. The Positive Predictive Power (PPP), Negative Predictive Power (NPP), specificity, and sensitivity of the CBCL

	PPP	NPP	Specificity	Sensitivity
Boys	59	96	92	74
Girls	36	97	81	80

Discriminative power of the CBCL

To examine if children with ADHD obtain different AP-scores than children without ADHD, mean AP-scores by informant were computed. The results are reported in Table 4. Main effects of sex, diagnosis, and their interaction were examined with a two-way ANOVA.

Table 4. Mean AP-scores of the total sample of boys and girls by diagnosis and correlations between AP and the number of ADHD symptoms

AP-score	N No ADHD		N ADHD		Mean AP score (SD) No ADHD		Mean AP score (SD) ADHD		Correlations	
	M	F	M	F	M	F	M	F	M	F
maternal, age 10	234	254	44	36	3.66 (3.44)	3.18 (3.37)	9.75 (4.24)	8.06 (3.83)	.63	.62
paternal, age 10	176	189	30	23	2.77 (2.74)	2.23 (2.71)	8.63 (4.26)	6.91 (3.55)	.64	.62
teacher, age 10	133	136	25	16	7.10 (7.19)	4.69 (6.04)	13.92 (7.69)	6.50 (6.45)	.32	.19
maternal, after interview	196	220	34	30	2.88 (2.93)	2.61 (2.81)	8.91 (4.62)	7.53 (4.22)	.69	.68
Total number	238	255	45	36						

M=Male, F=Female

Children with ADHD scored significantly higher than children without ADHD on both maternal and paternal AP-scales. The AP scores obtained from teachers did not discriminate significantly between groups of children with and without ADHD, although the mean scores of boys with ADHD were almost twice as large as the mean scores of boys without ADHD.

Regression analyses

To examine the predictive value of AP, linear stepwise regression analyses were performed. The dependent variable was the total number of ADHD symptoms. We included maternal AP scores collected after the interview, and paternal and teacher AP scores collected at age 10 years. The best fitting model in the first random sample was a model that includes both maternal and paternal AP scores plus an interaction effect between sex and the maternal AP score. This model explained 55% of the variance in the number of ADHD symptoms. The inclusion of teacher data did not improve the prediction of ADHD. The standardized regression weight of the maternal AP score was .47, of the paternal AP score .41, and of the interaction effect between sex and maternal AP -.13. The negative regression weight of the interaction effect implies that for the same values of AP, girls are expected to have a lower number of ADHD symptoms than boys. The results of the same regression analyses in the second random sample were very similar. However, no significant interaction effect between sex and AP was found.

DISCUSSION

We examined the relations between parental and teacher ratings on Attention Problems, and interview based DSM-ADHD diagnoses in a general population sample of boys and girls. Children were selected on the basis of their maternal CBCL scores at ages 7, 10, and 12 years. The mothers of these children were interviewed with a standardized DSM-interview to verify the presence of ADHD.

THE ASSOCIATION BETWEEN AP-SCORES AND DSM-ADHD

We examined the cross-sectional association between DSM-ADHD and maternal CBCL-AP, and found that CBCL-AP is an excellent screening instrument for the absence of ADHD in a population sample. However, screening for the presence of ADHD is associated with a high proportion of false positive cases. Of the total sample of children with a high maternal CBCL-score, 59% of the boys and 36% of the girls was diagnosed positive for ADHD. The remaining children were diagnosed negative for ADHD. This suggests that the CBCL can be used as a screening instrument for ADHD, but that children who score high on the CBCL have to be examined with additional methods to verify that they do indeed have ADHD.

The PPP was low compared to the PPP in most other studies (compare Table 1 and Table 3b). Because the PPP depends on the prevalence of the disorder, this can probably be explained by the fact that the prevalence of ADHD was much lower in our sample than the prevalence in clinically referred samples. The baseline prevalence of ADHD was 14% in boys and 12% in girls. In children with a high CBCL score, these percentages increased to 59% and 36%, respectively. In contrast, the baseline prevalence in studies of referred samples ranged from 29% to 83%. Because of these higher baseline rates, the PPP will be higher even if the association between the checklist and the interview is similar. The results of Gould, Bird, and Staghezza Jaramillo (1993), who studied a non-referred sample in Puerto Rico, are in agreement with our results. The prevalence of ADHD in their sample was 23%; this percentage increased to 36% conditional on a high CBCL score. This supports the hypothesis that the different levels of PPP are caused by different baseline prevalences. Concluding, selecting children with ADHD from a general population sample is associated with more false positive cases than selecting children with ADHD from a referred sample.

SEX DIFFERENCES IN THE ASSOCIATION BETWEEN AP AND ADHD

Sex differences were found in most association measures. Firstly, the PPP was higher in boys than in girls. This means that a boy with a high CBCL-score has a higher chance of obtaining a positive diagnosis for ADHD than a girl with a high CBCL-score. This is in agreement with the results of Chen et al. (1994), who also found higher positive predictive power in boys than girls. Secondly, the correlations between AP and the number of ADHD symptoms were higher in boys than girls, regardless of informant. Thirdly, regression analyses revealed significant negative interaction effects between AP and sex. The negative interaction effect implies that, for a certain value of AP, the predicted number of ADHD symptoms is lower in girls than it is in boys. This is in agreement with the lower predictive power in girls than boys. A possible explanation for the higher association between AP and ADHD in boys than girls is that symptoms that are displayed by boys with ADHD are more likely to be included in the DSM diagnosis than symptoms that are displayed by girls with ADHD. This could be caused by the fact that the development of ADHD was based on data from studies with more boys than girls, or because ADHD has a higher prevalence in boys than girls (Gaub & Carlson, 1997). In other words, the diagnosis of ADHD may be biased towards the male manifestation of the syndrome. If this hypothesis is correct, this may result in the underidentification of girls with ADHD, as has already been suggested by Hudziak (2000).

MULTI-INFORMANT CHECKLIST DATA

The correlations between paternal AP-scores and ADHD were just as high as the correlations between maternal AP-scores and ADHD. Although fathers report specific and unique aspects of their childrens behavior, we observed overall agreement between father reports on a checklist and mother reports in interview on ADHD behaviors.

An unexpected finding was that children with and without ADHD did not differ significantly with respect to their teacher ratings at age 10. The lower association between teacher reports and DSM-interviews may be due to a variety of factors, most prominently the fact that the DSM interview was collected from mothers and not teachers. Other factors include the different context of the children's behavior: teachers report on children's behavior in the classroom, parents report on children's behavior elsewhere. The finding of higher cross-correlations between paternal and maternal reports than between parental and teacher reports (Achenbach & Rescorla, 2001) is in line with this observation. The low discriminative power of teacher data was confirmed by low correlations between teacher ratings and the number of DSM-symptoms. The higher association between parental AP-ratings and ADHD than between teacher AP-ratings and ADHD may also reflect a better understanding of the child's behavior by the parents. The teachers had known the child only for an average of 12 months, while the parents had known the child since birth.

Regression analyses showed that the inclusion of paternal ratings slightly improves the prediction of ADHD. This suggests that a clinician should, whenever possible, obtain information from both parents when screening for the presence of behavior problems.

LIMITATIONS

This study concentrated on the association between behavior checklist scores and interview-based diagnoses. A limitation of this study is that the diagnosis for ADHD was only based on a DSM-interview with the mother. The results could have been different if the father, teacher or child was interviewed as well. However, the correlations between paternal AP-scores and ADHD were just as high as correlations between maternal AP-scores and ADHD.

Furthermore, we limited ourselves to the examination of the concurrent validity between two measures of ADHD, and did not examine the external validity of these measures. Previous research has shown that a combination of CBCL-scores and DSM-diagnosis provides a better prediction of outcome measures (e.g., disciplinary problems in school, and receiving inpatient or outpatient treatment) than either one of these instruments alone (Ferdinand et al., 2003). Finally, we made the choice to examine children in the age range of 10 to 13 years. Because of the exclusion of children below the age of 10 years or above the age of 13 years, we can not tell if our results generalize to children within these age groups.

REFERENCE LIST

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.

American Psychiatric Association (1994). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorder: DSM-IV (4th ed.)*. Washington, DC: American Psychiatric Association.

Biederman, J., Faraone, S. V., Doyle, A., Krifcher Lehman, B., Kraus, I., Perrin, J. et al. (1993). Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 34,* 1241-1251.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Bradley, J. D. D. & Golden, C. J. (2001). Biological contributions to the presentation and understanding of attention-deficit/hyperactivity disorder: a review. *Clinical Psychology Review, 6,* 907-929.

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the child behavior checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62,* 1017-1025.

de la Osa, N., Ezpeleta, L., Domenech, J. M., Navarro, J. B., & Losilla, J. M. (1997). Convergent and discriminant validity of the structured diagnostic interview for children and adolescents (DICA-R). *Psychology in Spain, 1,* 37-44.

Eiraldi, R. B., Power, T. J., Karustis, J. L., & Goldstein, S. G. (2000). Assessing ADHD and comorbid disorders in children: the Child Behavior Checklist and the Devereux Scales of Mental Disorders. *Journal of Clinical Child Psychology, 29,* 3-16.

Ferdinand, R. F., Heijmens Visser, J., Hoogerheide, K. N., Ende, J., Kasius, M. C., Koot, H. M. et al. (2003). Improving estimation of the prognosis of childhood psychopathology; combination of DSM-III-R/DISC diagnoses and CBCL scores. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44,* 1-10.

Ferdinand, R. F. & Ende van der, J. (1998). *Diagnostic Interview Schedule for Children IV Parent-version*. Rotterdam: Erasmus University Rotterdam, Department of Child and Adolescent Psychiatry.

Gaub, M. & Carlson, C. L. (1997). Gender differences in ADHD: a meta-analysis and critical review. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 1036-1045.

Gould, M. S., Bird, H., & Staghezza Jaramillo, B. (1993). Correspondence between statistically derived behavior problem syndromes and child psychiatric diagnoses in a community sample. *Journal of Abnormal Child Psychology, 21,* 287-313.

Hudziak, J. J. (2000). Genetics of attention-deficit/hyperactivity disorder. In T.E.Brown (Ed.), *Attention-deficit disorders and comorbidities in children, adolescents, and adults* (pp. 57-78). Washington, DC: American Psychiatric Press.

Hudziak, J. J., Copeland, W., Stanger, C., & Wadsworth, M. (2004). Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 1299-1307.

Jensen, P. S. & Watanabe, H. (1998). Sherlock Holmes and child psychopathology assessment approaches. The case of the false-positive. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 138-146.

Kasius, M. C., Ferdinand, R. F., Berg van den, H., & Verhulst, F. C. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38,* 625-32.

Kazdin, A. E. & Heidish, I. E. (1984). Convergence of clinically derived diagnoses and parent checklists among inpatient children. *Journal of Abnormal Child Psychology, 12,* 421-436.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, Section A, 62, Part 1,* 28-30.

Lengua, L. J., Sadowski, C. A., Friedrich, W. N., & Fisher, J. (2001). Rationally and empirically derived dimensions of children's symptomatology: expert ratings and confirmatory factor analyses of the CBCL. *Journal of Consulting and Clinical Psychology, 69,* 683-698.

Shaffer, D., Schwabstone, M., Fisher, P., Cohen, P., Piacentini, J., Davies, M. et al. (1993). The Diagnostic Interview Schedule for Children Revised Version (Disc-R) .1. Preparation, Field Testing, Interrater Reliability, and Acceptability. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 643-650.

Sprafkin, J., Gadow, K. D., Salisbury, H., Schneider, J., & Loney, J. (2002). Further evidence of reliability and validity of the Child Symptom Inventory-4: Parent Checklist in clinically referred boys. *Journal of Clinical Child and Adolescent Psychology, 31,* 513-524.

SPSS (2001). *SPS for Windows, Rel 11.0.1*. Chicago, Ill: SPSS.

Steinhausen, H. C., Winkler Metzke, C., Meier, M., & Kannenberg, R. (1997). Behavioral and emotional problems reported by parents for ages 6 to 17 in a Swiss epidemiological study. *European Child and Adolescent Psychiatry, 6,* 136-141.

Verhulst, F. C., Ende van der, J., & Koot, H. M. (1996). *Handleiding voor de CBCL/4-18*. Rotterdam: Afdeling Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.

CHAPTER

5

PROBLEM BEHAVIOR AT SCHOOL: WHY MORE BOYS THAN GIRLS ARE REFERRED FOR TREATMENT OF ADHD

This paper is submitted as Eske M. Derks, James J. Hudziak, Dorret I. Boomsma. Problem behavior at school: Why more boys than girls are referred for treatment of ADHD. *Comprehensive Psychiatry, submitted.*

ABSTRACT

More boys than girls with attention deficit hyperactivity disorder (ADHD) are referred for treatment. One explanation for this difference may be that boys score higher on disruptive behavior scales than girls. Although this was supported by findings in clinical samples, recent studies in non-referred samples showed that boys and girls with ADHD are similar with respect to their levels of disruptive behavior as reported by their mother. In this report, we investigate whether the difference in referral rate can be explained by higher teacher problem scores in boys with ADHD than in girls with ADHD. Data were obtained from mothers and teachers in a non-referred sample of 283 boys and 291 girls with and without ADHD. Children were selected if they scored either low (controls) or high (probands) on attention problems. Mothers completed DSM-IV interviews, Child Behavior Checklists (CBCL) and the Conners Rating Scale (CRS). Teacher completed the Teacher Report Form (TRF), and the CRS. Boys and girls with ADHD had similar levels of psychiatric illness and school impairment (such as being held back, special class placement, and learning problems) by mother report. Mothers reported similar levels of aggression and attention problems in boys and girls with ADHD. In contrast, teachers consistently rated boys with ADHD as having higher scores on reports of attention problems and ADHD related behavior than girls with ADHD. Gender differences vary across settings: boys and girls with ADHD are rated as behaving differently at school, but not at home. The higher level of teacher reported problem behavior at school may explain the high male:female ratio for ADHD in clinical settings. Furthermore, our results support the idea that the reliance on teacher reports in referral results in under identification of girls with ADHD.

INTRODUCTION

The boy:girl ratio for attention deficit hyperactivity disorder (ADHD) is estimated at 9:1 in clinical settings and 3:1 in the general population (Gaub & Carlson, 1997). It is unclear why so many more boys than girls are referred for treatment. A variety of hypotheses have been offered for this disparity. Hudziak (1997) suggested that the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria for ADHD were developed primarily in male samples and are therefore sensitive to the sex of the child. The application of diagnostic criteria based on male samples may lead to under identification of girls due to sex specific differences in symptom levels. Others have suggested that under identification may be due to the fact that girls have lower rates of comorbidity (Biederman et al., 2002).

In a review of the relations between ADHD and comorbid conditions, boys with ADHD showed higher rates of oppositional defiant disorder (ODD), conduct disorder (CD) and other externalizing problems than girls with ADHD (Gaub & Carlson, 1997). This finding has been replicated by Biederman et al. (2002) who found higher rates of comorbidity with depression, CD, and ODD in boys than in girls with ADHD. However, the interpretation of these findings is influenced by ascertainment factors, particularly the fact that the sample was clinically recruited. Indeed, in general population samples, no gender differences on comorbidity with CD and ODD have been found (Biederman et al., 2005; Levy et al., 2005). For internalizing disorders, comorbidity does vary by gender, but in this case girls with the "inattentive subtype" had higher rates of separation anxiety disorder while girls with the "combined subtype" showed a higher rate of generalized anxiety disorder (Levy et al., 2005).

A second explanation for the gender difference in referral rate may be a differential impact of ADHD symptoms in school settings. Is it possible that teachers report ADHD symptoms differently in boys and girls? Might it be that teacher reports account for the differences in referral rates? Much has been written about the role of school teachers in the identification of children in need of treatment for ADHD, but to our knowledge, only a few studies have specifically quantified gender differences in teacher ratings with respect to referral for treatment of ADHD. Several groups have studied the behavior of children with ADHD at school. In preschool, elementary and secondary school, nonreferred boys showed higher rates of ADHD than nonreferred girls (Nolan et al., 2001). In all age groups, both boys and girls with ADHD received higher scores on ODD and CD than children without ADHD, but the question whether the comorbidity with ODD and CD was higher in boys than girls with ADHD was not addressed. A significant interaction of gender by diagnosis was found in teacher report data on children with ADHD as identified in a school population (Carlson et al., 1997). Girls with the ADHD-combined type received lower scores on aggressive (AGG) and externalizing behavior than boys with the ADHD combined type. However, these two studies are based on data from DSM-IV Symptom checklists obtained from teachers, and not on clinical diagnoses. Using the Diagnostic Interview Schedule for Children (DISC-IV) with mothers to validate the presence or absence of ADHD, Abikoff et al. (2002) used trained observers who were blind to diagnostic status to rate the behavior of 502 children with ADHD and paired comparisons (i.e., children without ADHD) in the classroom. A DSM-interview with the mother was used to validate the presence or absence of ADHD. The authors report that girls with ADHD obtained lower scores on behaviors related to ADHD and AGG than boys with ADHD. In summary, teachers report more disruptive behaviors in boys than girls at school. It may be that this gender specific finding, in concert with the ADHD diagnosis, leads to the difference in referral rates between boys and girls.

In the present study, our aim is to determine how gender, comorbidity, and informant effects influence a child's access to treatment for ADHD. Data were collected in a sample of Dutch boys and girls selected based on scores on the attention problems (AP)

syndrome of the CBCL. A structured clinical interview was completed by mothers to determine if a child had ADHD and to assess for psychiatric comorbidity. Mothers and teachers completed behavior checklists on AP and aggression (AGG). Finally, we collected data on school impairment (e.g., repeated class, special class placement, learning problems), and treatment rate. We addressed the following questions: Are gender differences in treatment referral explained by i) higher levels of psychiatric comorbidity; ii) higher levels of behavioral disturbance at home; iii) higher levels of behavioral disturbance at school; or iv) greater school impairment in boys than girls.

METHODS

SUBJECTS

The subjects are Dutch twins whose parents voluntarily registered with the Netherlands Twin Registry when the twins were born (Boomsma et al., 2002). As described previously (Derks et al., 2006) , subjects were selected on the basis of standardized maternal Child Behavior Checklist (CBCL) (Achenbach, 1991a) ratings collected at the ages 7, 10, and 12 years. More specifically, twin-pairs were selected if at least one of the twins scored high on AP (probands), or if both twins scored low on AP (controls). Standardized T-scores (Mean=50; SD=10) were computed within gender. A high score was defined as a T-score above 60 at all available time-points (age 7, 10, and 12 years) and a T-score above 65 at least once. A low score was defined as a T-score below 55 at all time-points. The control twins were matched with proband twins on the basis of gender, cohort, maternal age, and social economic status. The resulting sample consisted of 283 males and 291 females. At the time of the interview, the twins had a mean age of 11.50 years (SD=.68). After complete description of the study to the families, all parents signed written consent for participation.

INSTRUMENTS

The mothers were interviewed by two experienced research assistants. Psychiatric assessments were made with the Diagnostic Interview Schedule for Children (DISC-IV) (Shaffer et al., 1993). A child was diagnosed positive for ADHD if he or she met type-A criteria of the DSM-IV (American Psychiatric Association, 1994). Two-hundred-forty-eight interviews were audiotaped. A research assistant, who was blind to the results of the interview, listened to 40 of these interviews. The number of symptoms scored by the interviewer and the research assistant showed perfect agreement.

Mothers completed a CBCL within four months after the interview (Achenbach, 1991a), and a Conners' Parent Rating Scale (Conners et al., 1998b) when the children were twelve years old. Teachers completed the Teacher Report Form (TRF) (Achenbach, 1991b) and the Conners' Teacher Rating Scale (Conners et al., 1998a) when the children were twelve years old. In the statistical analyses, we included the attention problem scale (11 items) and aggression scale (20 items) of the CBCL, the attention problem scale (20 items) and aggression scale (25 items) of the TRF, and the ADHD-index (ADHD-I; 12 items) of the maternal and teacher version of the Conners' Rating Scale.

Treatment was assessed by means of two questions that are included within the DISC. The question that was used to assess medication use is "In the past 12 months, did he/she use medication for overactivity, hyperactivity or attention problems?" The question that was used to assess clinical referral is "In the past twelve months, did he/she visit someone in a hospital, outpatient clinic or other institution because he/she was overactive or hyperactive or had attention problems?". Three measures of maternal reports on school impairment were included: i) did the child ever repeat a grade?; ii) was the child placed in a special class?; and iii) did mother report learning problems in the surveys collected when the children were 7 or 10 years old?

Statistical analyses

Chi square tests (χ^2) were performed to assess the effects of gender and diagnosis on psychiatric comorbidity, school impairment and treatment rate in the total sample. In order to determine which variables contribute to the different levels of referral rate in boys and girls, the effect of gender was also examined within the group of children with ADHD. Students t-tests were used to examine the effect of gender and diagnosis on problem behavior scores as reported by teachers and parents in the total population and the effect of gender in the population of children with ADHD. Statistical significance of the two-tailed tests was determined using $\alpha<.05$.

RESULTS

Demographics

There were 45 boys and 36 girls with ADHD. The relative frequencies of the subtypes were significantly different in boys and girls (χ^2 (3) = 10.5, p = .01). In boys, the combined type (CT) was most common (N=22; 49%), followed by the inattentive type (IN) (N=14; 31%), and the hyperactive-impulsive type (HI) (N=9; 20%). In girls, IN had the highest prevalence (N=18; 50%), followed by HI (N=12; 33%) and CT (N=6; 17%).

Treatment rate

The number of children who receive treatment for problems related to ADHD are summarized by gender and ADHD status in Table 1. As expected, children with ADHD more often received medication (χ2(1)=89, p<.001) and counseling (χ^2(1)=62, p<.001) for ADHD related problems than children without ADHD. As can be seen in Table 1, in the total sample of children with and without ADHD, boys more often received medication and counseling than girls. Likewise, boys with ADHD more often received medication and counseling than girls with ADHD. This higher referral rate in boys was not due to gender differences in the rates of ADHD subtypes. For example, in children with ADHD-CT, 64% of the boys and 17% of the girls received medication.

Psychiatric Comorbidity, school Impairment, and behavior problem scores in children with and without ADHD

Psychiatric comorbidity, school impairment, and CBCL, TRF, and CRS scores were compared between children who meet DSM-IV criteria for ADHD and children who do not meet these criteria. The results are summarized in Table 2. In all areas, children with ADHD performed worse than children without ADHD. ADHD children had higher prevalences of oppositional defiant disorder, conduct disorder, generalized anxiety disorder, and major depressive disorder. However, no significant difference was found for separation anxiety disorder. Furthermore, ADHD children were more impaired at school, and obtained higher scores on maternal and teacher checklist ratings of AP, AGG, and ADHD.

Table 1. Prevalence and treatment rates of attention deficit hyperactivity disorder (ADHD) by subtype in male and female twins with and without ADHD

	Medication					Counseling				
	Boys		Girls		Effect of gender	Boys		Girls		Effect of gender
	N	%	N	%	χ^2 (df)	N	%	N	%	χ^2 (df)
Total sample	30	11	3	1	24(1)***	27	10	6	2	15(1)***
No ADHD	9	4	1	0	7(1)**	10	4	3	1	4 (1)*
ADHD	21	47	2	6	17(1)***	17	38	3	8	9 (1)**

***p<.001; **p<.01; *p<.05

Table 1. Prevalence and treatment rates of attention deficit hyperactivity disorder (ADHD) by subtype in male and female twins with and without ADHD *(Continued)*

	Medication					Counseling				
	Boys		Girls		Effect of gender	Boys		Girls		Effect of gender
	N	%	N	%	χ^2 (df)	N	%	N	%	χ^2 (df)
By subtype										
Inattentive	3	21	1	6	2(1)	5	36	1	6	5 (1)*
Hyperactive-Impulsive	4	44	0	0	7(1)*	3	33	0	0	5 (1)*
Combined	14	64	1	17	4(1)*	9	41	2	33	0 (1)

***p<.001; **p<.01; *p<.05

Table 2. Psychiatric comorbidity, treatment, school impairment, and maternal and teacher reports of problem behavior in male and female twins with and without attention deficit hyperactivity disorder (ADHD)

	ADHD				No ADHD				Total group		ADHD group
	Boys (N=45)		Girls (N=36)		Boys (N=238)		Girls (N=255)		Gender	ADHD status	Gender
Psychiatric comorbidity	N	%	N	%	N	%	N	%	χ^2 (df)	χ^2 (df)	χ^2 (df)
Oppositional defiant disorder	13	29	10	28	7	3	9	4	0(1)	70(1)***	0(1)
Conduct disorder[a]	0	0	2	6	0	0	0	0	-	-	-
Generalized anxiety disorder	5	11	3	8	0	0	3	1	0(1)	32(1)***	0(1)
Separation anxiety disorder	2	4	3	8	3	1	6	2	1(1)	6(1)*	1(1)
social phobia	8	18	2	6	2	1	8	3	0(1)	22(1)***	3(1)
Specific phobia	10	22	9	25	18	8	22	9	0(1)	18(1)***	0(1)
Depression	2	4	4	11	1	0	2	1	1(1)	21(1)***	1(1)
School impairment	N	%	N	%	N	%	N	%	χ^2(df)	χ^2(df)	χ^2(df)
repeated grade	18	41	15	42	54	23	58	23	0(1)	13(1)***	0(1)
special class	10	22	3	8	12	5	9	4	3(1)	17(1)***	3(1)
Learning problems	26	58	24	67	78	33	79	31	0(1)	27(1)***	1(1)
Maternal reports	Mean	SD	Mean	SD	Mean	SD	Mean	SD	t(df)	t(df)	t(df)
attention problems (N=208)	8.9	4.6	7.5	4.2	3.0	3.0	2.7	2.8	2(508)	10(71)*** [b]	1(62)
aggression (N=208)	15.9	8.6	12.6	8.0	5.8	5.5	4.6	4.6	3(508)**	9(70)***[b]	2(62)
ADHD-index (N=100)	21.8	8.4	16.4	7.2	7.9	7.2	6.3	6.8	2(218)*	9(218)***	2(27)
Teacher reports	Mean	SD	Mean	SD	Mean	SD	Mean	SD	t(df)	t(df)	t(df)
attention problems (N=92)	14.3	7.1	6.6	7.2	4.8	5.2	2.9	4.3	4(221)***	5(39)***[b]	3(33)**
aggression (N=92)	12.4	10.9	5.9	8.8	4.2	5.8	2.2	4.0	4(221)**	4(37)**[b]	2(33)
ADHD-index(N=91)	14.1	8.2	7.2	7.6	4.3	5.6	2.1	4.0	4(216)***	5(36)***[b]	2(31)*

[a] No statistical test could be performed due to the small number of children who met criteria for conduct disorder
[b] Levene's test revealed a significant difference in variance, therefore statistical tests were performed without the assumption of equality of variances
***p<.001; **p<.01; *p<.05

Gender differences in the total population

The prevalence of psychiatric disorders was not different in boys and girls (Table 2). Likewise, boys and girls showed similar levels of school impairment. Gender differences were observed in behavior checklist ratings. As expected, in the general population sample, boys had higher scores on maternal ratings of AGG and the ADHD-I, and on teacher ratings on AP, AGG, and the ADHD-I.

Gender differences in children with ADHD

In children with DSM-IV ADHD, there were no gender differences in comorbidity profiles or on measures of school impairment (Table 2). Maternal CBCL and Conner's checklist scores also were not significantly different between boys and girls with ADHD. However, in teacher ratings, some interesting differences emerged. Boys obtained higher scores on teacher ratings of AP, and on the teacher rated ADHD-I. Teacher ratings on AGG were twice as high in boys with ADHD as in girls with ADHD, but this difference was not statistically significant.

DISCUSSION

The purpose of the present study was to investigate why boys with ADHD are more often referred for treatment than girls with ADHD. Psychiatric comorbidity, school impairment, and problem behavior at home and at school were assessed in a sample of boys and girls who were selected on the basis of maternal AP scores. Most studies on gender differences in children with ADHD were conducted in clinical samples. This study adds to the relatively rare results on gender differences in nonreferred samples of children with ADHD.

It appears that girls with ADHD have similar profiles of psychiatric comorbidity, have similar levels of school impairment, but are far less likely to receive treatment than boys with ADHD. Of the girls with ADHD, only 6% was prescribed medication and 8% received counseling, compared to 47% and 38%, respectively, in boys. These results indicate that ADHD is under treated in girls relative to boys, and also, that the majority of both boys and girls with ADHD are not being treated. Our findings are consistent with those of Reich, Huang & Todd (2006) who studied ADHD medication use in a large sample of boys and girls from the state Missouri in the US. They showed that for children who meet full criteria of DSM-IV ADHD, 75% of the boys and 68% of the girls were in treatment. In addition, 59% of the boys and 46% of the girls received medication. A third important finding is a gender-by-country bias, as a far greater percentage of girls with ADHD are receiving treatment in the US samples as compared to the girls with ADHD in our Dutch sample. It is fairly clear that a referral bias exists that identifies far less girls than boys for treatment than would be expected. This is in line with the observation that the gender ratio is 1:3 in general population samples compared to 1:9 in clinical samples (Gaub & Carlson, 1997). In the current study, we investigated a number of factors that may contribute to this referral bias.

Consistent with the results of Biederman et al. (2005), we found that boys and girls with ADHD obtained similar rates of disruptive behavior disorders. Likewise, continuous measures of ADHD and AGG showed that boys and girls with ADHD have similar levels of problem behavior at home: mothers report fairly similar levels of problem behaviors in boys and girls with ADHD. In contrast, *teacher* reports of ADHD and AGG do discriminate between boys and girls with ADHD. Teachers report lower levels of problem behavior in girls with ADHD than in boys with ADHD. Although the source of this discrepancy is not clear, the difference between teacher and mother reports on girls is a clear point of disagreement between adults, who contribute information that is relevant to the decision to refer a child for treatment. In the case of boys, where mother reports high levels of ADHD, inattention, and AGG, the teacher reports often concur, thus resulting in relative agreement about the need to refer for treatment. In the case of girls, however, although mother may recog-

nize and report high levels of ADHD, inattention and AGG at home, the teacher reports often do not support the same high levels ADHD symptoms, thus diminishing the consensus for the need for referral. These data support the contention of many that teacher reports provide the main impetus for referring a young child for treatment. However, the implication of the present results is that, while the teachers are correctly identifying boys with ADHD, they are under identifying girls with ADHD. Therefore, the teacher rating effect may explain the referral gender-bias in ADHD in pediatric clinic settings.

An alternative possibility is that the referral bias is due to a different manifestation of the disorder in boys and girls. For example, girls with ADHD may have better physical or psychosocial health than boys with ADHD. Klassen, Miller & Fine (2004) studied health-related quality in life in boys and girls with ADHD and showed a slightly poorer physical health in boys with ADHD than in girls with ADHD. No gender difference was reported for psychosocial health. In two different areas, however, girls have been shown to perform worse than boys. First, a follow-up study of boys and girls with ADHD revealed a higher risk of adult psychiatric admission in girls than boys (Dalsgaard et al., 2002). In this study, 208 subjects (183 boys and 25 girls), who were referred for hyperactivity/inattention and received treatment with stimulants between 1969 and 1989, were identified. Twenty-three percent of the subjects had a psychiatric admission in adulthood; this rate was higher in females than in males. Second, girls with ADHD showed more social problems than boys with ADHD. For example, they were more likely to suffer peer rejection than boys with ADHD (Berry et al., 1985). Furthermore, in the present study, we found similar levels of school impairment in boys and girls with ADHD. In conclusion, it is unlikely that the lower referral rate in girls can be explained by a better disease prognosis.

The toll in terms of morbidity and mortality of the lower treatment rates in girls may be considerable. The long term effect of untreated ADHD is well documented. Increased risk of psychiatric disorders (e.g., major psychopathology, anxiety disorders, antisocial disorders, developmental disorders, and substance dependence disorders) in young adulthood (Biederman et al., 2006) and of poor adult achievement (Barkley, 2002) may all be ameliorated, provided girls with ADHD are correctly identified and treated. One rather simple approach to correcting the gender disparity is to simply rely on parent reports to identify girls, who should be screened for having ADHD. In boys, pediatricians may want to base the diagnosis on reports of behavior problems in both the home and the school environment. For girls, it may be sufficient to establish that they are displaying ADHD related problems at home. Results of the present and other studies indicate that, if girls are described by their mothers as having ADHD, they are likely to meet diagnosic criteria, have similar comorbidity profiles as boys with ADHD, and are academically impaired. Relying on secondary teacher reports to confirm the diagnoses of ADHD in girls may lead to the under identification of girls suffering from this serious medical condition.

Conclusion

Overall, our study calls for a careful approach to the assessment of emotional and behavioral problems in girls. In this study we demonstrated a robust informant by gender interaction that may lead to under identification in girls of one of the most common and impairing child psychopathologies. Furthermore we combined our results with those of others to provide support for the presence of gender bias in both the US and in the Netherlands, and for the role of cultural factors (with even fewer girls with ADHD receiving treatment in the Netherlands than in the US). While we suggest that diagnosis in girls should be based on parental informants, it should be realized that any single rater (e.g., mother) strategy is also potentially biased. Therefore, sex specific approaches towards the assessment of ADHD should be developed.

LIMITATIONS

In interpreting the results of the present study, the following limitations should be taken into account. First, further study is required to establish whether the present Dutch population results generalize to population samples in the United States. Second, clinical diagnoses were based on structured diagnostic interviews with the mother. The results may be different when the assessment of ADHD is based on expert clinical diagnoses.

REFERENCE LIST

Abikoff, H. B., Jensen, P. S., Arnold, L. L. E., Hoza, B., Hechtman, L., Pollack, S. et al. (2002). Observed classroom behavior of children with ADHD: Relationship to gender and comorbidity. *Journal of Abnormal Child Psychology, 30,* 349-359.

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

American Psychiatric Association (1994). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorder: DSM-IV (4th ed.)*. Washington, DC: American Psychiatric Association.

Barkley, R. A. (2002). Major life activity and health outcomes associated with attention-deficit/hyperactivity disorder. *Journal of Clinical Psychiatry, 63,* 10-15.

Berry, C. A., Shaywitz, S. E., & Shaywitz, B. A. (1985). Girls with Attention Deficit Disorder - A Silent Minority - A Report on Behavioral and Cognitive Characteristics. *Pediatrics, 76,* 801-809.

Biederman, J., Kwon, A., Aleardi, M., Chouinard, V. A., Marino, T., Cole, H. et al. (2005). Absence of gender effects on attention deficit hyperactivity disorder: Findings in nonreferred subjects. *American Journal of Psychiatry, 162,* 1083-1089.

Biederman, J., Mick, E., Faraone, S. V., Braaten, E., Doyle, A., Spencer, T. et al. (2002). Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. *American Journal of Psychiatry, 159,* 36-42.

Biederman, J., Monuteaux, M. C., Mick, E., Spencer, T., Wilens, T. E., Silva, J. M. et al. (2006). Young adult outcome of attention deficit hyperactivity disorder: a controlled 10-year follow-up study. *Psychological Medicine, 36,* 167-179.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Carlson, C. L., Tamm, L., & Gaub, M. (1997). Gender differences in children with ADHD, ODD, and co-occurring ADHD/ODD identified in a school population. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 1706-1714.

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998a). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26,* 279-291.

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998b). The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26,* 257-268.

Dalsgaard, S., Mortensen, P. B., Frydenberg, M., & Thomsen, P. H. (2002). Conduct problems, gender and adult psychiatric outcome of children with attention-deficit hyperactivity disorder. *British Journal of Psychiatry, 181,* 416-421.

Derks, E. M., Hudziak, J. J., Dolan, C. V., Ferdinand, R. F., & Boomsma, D. I. (2006). The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. *Comprehensive Psychiatry, 47,* 116-122.

Gaub, M. & Carlson, C. L. (1997). Gender differences in ADHD: a meta-analysis and critical review. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 1036-1045.

Hudziak, J. J. (1997). Identification of phenotypes for molecular genetic studies of common childhood psychopathology. In K.Blum & E. Noble (Eds.), *Handbook of Psychiatric Genetics* (pp. 201-217). Boca Raton: CRC Press.

Klassen, A. F., Miller, A., & Fine, S. (2004). Health-related quality of life in children and adolescents who have a diagnosis of attention-deficit/hyperactivity disorder. *Pediatrics, 114,* E541-E547.

Levy, F., Hay, D. A., Bennett, K. S., & McStephen, M. (2005). Gender differences in ADHD subtype comorbidity. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 368-376.

Nolan, E. E., Gadow, K. D., & Sprafkin, J. (2001). Teacher reports of DSM-IV ADHD, ODD, and CD symptoms in schoolchildren. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 241-249.

Reich, W., Huang, H., & Todd, R. D. ADHD medication use in a population-based sample of twins. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*, 801-807.

Shaffer, D., Schwabstone, M., Fisher, P., Cohen, P., Piacentini, J., Davies, M. et al. (1993). The Diagnostic Interview Schedule for Children Revised Version (Disc-R) .1. Preparation, Field Testing, Interrater Reliability, and Acceptability. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 643-650.

CHAPTER

6

STATISTICAL POWER TO DETECT GENETIC AND ENVIRONMENTAL INFLUENCES IN THE PRESENCE OF DATA MISSING AT RANDOM

This paper is in press as: Eske M. Derks, Conor V. Dolan, Dorret I. Boomsma. Statistical power to detect genetic and environmental influences in the presence of data missing at random. *Twin Research and Human Genetics, in press.*

ABSTRACT

The goal of the current simulation study is to investigate the optimal selection design for the situation in which a cheap measure (X) is observed in a large, representative twin sample, and a more expensive measure (Y) is observed in a selected subsample of the twins. Scores obtained on X are used to select the most informative twin-pairs with respect to the detection of genetic and environmental influences on the covariance between X and Y and on the variance of Y. Missingness (88%) was introduced in accordance with six selection designs: i) design based on an empirical study of the Netherlands Twin Registry; ii) extreme concordant design; iii) extreme concordant and discordant design (EDAC); iv) extreme discordant design; v) individual selection design; and vi) missing completely at random. In simulation study 1, we investigated the statistical power on the covariance structure in MZ and DZ twins as a function of the phenotypic correlation between X and Y. In simulation study 2, we investigated the statistical power to detect additive (A) and dominant (D) genetic effects, and shared environmental effects (C) on the covariance between X and Y and on the variance of Y in the six selection designs. The selection of informative pairs is most effective when the correlation between traits X and Y is high. The best selection design in terms of the statistical power to detect genetic and environmental influences on the variance of Y and on the covariance between X and Y is the individual selection design. The power to detect A and C is acceptable in most other designs as well. The power to detect D is low irrespective of selection design. If the screening variable is included as the X variable in a bivariate Full Information Maximum Likelihood analysis, the parameter estimates are unbiased. The decrease in statistical power as a result of missingness is relatively small. The individual selection design is the best design for detecting influences of A, C, D, and E but the EDAC design is preferred when an additional purpose of a study is to detect QTL effects.

INTRODUCTION

Questionnaires can be a cost-effective way to obtain information on a wide variety of phenotypes, (e.g., behavior, health, and social environment). Because of the relatively low costs and the possibility of mailing questionnaires, it is feasible to phenotype large numbers of subjects. Therefore, twin registries often include survey data collected using questionnaires in twins and their family members. For some purposes, however, it may be necessary to collect more expensive phenotypic (e.g., endophenotypic or biological) measures. For example, in gene finding studies, high costs are involved in the collection of DNA samples and the subsequent genotyping. Endophenotyping may include endocrine measures, assessment of neurocognitive measures, or assessment of brain structure volumes and functioning with Magnetic Resonance Imaging (MRI). Expensive phenotypes are sometimes collected through psychiatric interviews or 24 hour ambulatory recordings of cardiovascular functions. Because of the high costs associated with such measures, the number of subjects that can be tested is often limited. This may force one to phenotype selected subjects from those present in the large representative sample. Such selection should be optimized by selecting the most informative cases given the objective of the study.

The purpose of the present paper is to investigate the precision of parameter estimates in various selection designs in the context of multivariate genetic covariance structure modeling of MZ and DZ twin data (Martin & Eaves, 1977). Specifically we envisage the situation in which we want to estimate the genetic and environmental covariance structures of phenotypes X and Y. Relatively cheap measures of phenotype X are available in a large representative sample of twin-pairs. Phenotype Y, in contrast, is expensive to measure, and

can only be measured in a relatively small selected subsample. The question is: how should we select cases from the representative samples so that we retain the greatest possible statistical power, while ensuring that our estimates are unbiased?

Sibpair selection in Quantitative Trait Loci (QTL) analyses (Risch & Zhang, 1995; Gu et al., 1996; Eaves & Meyer, 1994) is designed to enrich the sample with sibs who share zero or two alleles identically by descent. The present objective is different, namely to obtain estimates of the genetic and environmental contributions to the variance of X and Y, and to the covariance between X and Y. To achieve this aim given the present limitations (i.e., limited means to phenotype Y), it would seem rational to select individuals who score extremely high or low on X. Provided that the correlation between X and Y is greater than zero, subjects with extreme scores on X will also have, on average, more extreme scores on Y. The linear relationship of X on Y can be estimated well with extreme high and low values on Y, as in combination these determine the orientation of the regression line in the regression of X on Y. The loss of power to detect a correlation between X and Y, as a result of missingness, can be very small indeed in this situation (Dolan et al., 2005). The present case is more complicated, as we have twin-pairs and therefore four rather than two variables. Although it may be expected that the selection of subjects with extreme scores on X will provide the best statistical power, a number of questions can be raised with respect to the selection design. Should we select concordant and/or discordant twin-pairs? Should we select on a twin-pair or on an individual basis? What is the effect of the phenotypic correlation between X and Y? To answer these questions we shall consider a number of selection designs.

The introduction of missingness due to selection does not pose an estimation problem, as full information (raw data) maximum likelihood (FIML) estimation can be used if the data are missing at random (MAR) or missing completely at random (MCAR), in the sense of Rubin (1976) (see also Little & Rubin, 2002; Schafer & Graham, 2002). We briefly explain the concepts of MCAR and MAR. The distributions for the missingness (R) can be classified according to the nature of the relationship between the missingness and the data (Schafer & Graham, 2002). Let the complete data matrix Z_{com}, be partitioned as $Z_{com}=(Z_{obs}, Z_{mis})$, where Z_{obs} and Z_{mis} are the observed and missing parts, respectively. Missing data are MAR if the distribution of missingness does not depend on Z_{mis},

$$p(R|Z_{com})=p(R|Z_{obs}). \quad (1)$$

In other words, the probability of missingness depends only on the observed part of the data, and not on the missing part. For example, suppose that blood pressure was observed in a sample of 1000 subjects, and that some additional data was collected in all subjects who obtained a blood pressure score of one standard deviation (sd) above average. Here, the probability of missingness is 1, given that a subject scores below the cutoff, and the probability of missingness is 0, given that a subject scores above the cutoff. The data are MAR, because the missingness depends on observed data only. This example is relevant to the present undertaking, here we have observed X, and on the basis of X we select twin-pairs for phenotyping with respect to Y.

A special case of MAR is missingness completely at random (MCAR). In the previous example, the data would be MCAR if a random sample of the 1000 subjects is invited to participate, and the probability of non-participation is not related to the trait of interest. In other words, the distribution of missingness does not depend on Z_{mis} or Z_{obs}.

$$p(R|Z_{com})=p(R) \quad (2)$$

When data are MAR or MCAR, a number of methods are available to deal with the missingness. These involve multivariate analysis of all available data with Full Information Maximum Likelihood (FIML) estimation, imputation of the missing data (Little & Rubin, 2002) and data weighting (Little & Rubin, 2002; Heath et al., 1998).

In the present paper, we apply FIML in six selection designs. In five of these designs, selection gives rise to data MAR; in the sixth design, which we include for reasons of comparison, the selected data are MCAR. First, we present the results of some preliminary analyses to investigate the effect of the selection design and the phenotypic correlation between X and Y on the estimates of the twin covariances. Next, we present the results of actual genetic covariance structure modeling using simulated twin data, with missingness due to selection according to six designs. We simulated data according to two different aetiological models, and fitted the models using FIML estimation in Mx (Neale et al., 2003). We report means and standard deviations (interpretable as standard errors) of the estimates of genetic and environmental parameters in each design. We also consider the power to detect additive genetic, dominant genetic, and shared environmental effects.

METHODS

A DESCRIPTION OF THE SIX SELECTION DESIGNS

Data are MAR in selection designs 1-5, and are MCAR in selection design 6. Designs 1-5 are based on the selection of twins, who score lowly or highly on the (cheap) phenotypic test X.

The first design considered here is a selection design that was used in a study on Attention Problems (AP) and attention deficit hyperactivity disorder (ADHD) (Derks et al., 2006). Twin-pairs were selected if: i) both members obtained low AP checklist scores, and ii) at least one of the members obtained a high AP checklist score. The lower threshold was not very extreme, but only 25% of the twin-pairs that obtained a low AP score was randomly selected. In selection design 2 (Extremely Concordant; EC), twin-pairs are selected if both sibs scored extremely high, and if both sibs scored extremely low. Selection design 3 is the EDAC design. In this design, pairs are selected if the two members of a twin-pair are discordant, and if both members score extremely low or high. In selection design 4, only extremely discordant pairs (ED) are selected. In selection design 5, twins are selected on an individual basis rather than on a pair-wise basis. Specifically, a complete twin-pair is selected if at least one of the twin scores extremely low or high. Finally, in selection design 6 (MCAR), twin-pairs are selected completely at random. In addition to analyzing the data obtained in these 6 selection designs, we also analyzed the complete dataset (i.e., no selection, and therefore no missingness).

DATA SIMULATION

All data were simulated using routines in the freely available R program (Venables et al., 2002). The data were multivariate normally distributed, with unit variances and zero means. In most European countries, the number of DZ twins is about twice the number of MZ twins. Therefore, we obtained a 1:2 ratio for the number of MZ:DZ twins. The number of MZ twin-pairs was 2000, and the number of DZ twin-pairs was 4000. The number of replications was 100. In the simulated data, 88% missingness was introduced in accordance with the selection design. To establish a constant percentage of missingness of 88%, the thresholds were allowed to vary in the six selection designs. Because of the different twin covariances in monozygotic (MZ) and dizygotic (DZ) twins, the percentages of missingness may vary between these groups. For example, given that phenotype X is a heritable trait, the twin correlations will be lower in DZ than in MZ twins. As a result, more DZ than MZ twins will be selected in the extreme discordant design. The choice of the thresholds and the effect sizes (i.e., the contributions of the genetic and environmental factors to the phenotypic covariance) are provided in the results section.

Simulation study 1: Estimation of MZ and DZ covariances in the presence of missingness

For each of the six selection designs, data were simulated according to two population covariance matrices in MZ and DZ twins. In the first covariance matrix, the phenotypic correlation between X and Y was high (r=.8), in the second, the phenotypic correlation was moderate (r=.4). The MZ and DZ correlations for both X and Y were 0.7 and 0.3, respectively. The order of the variables in the covariance matrices are: variable X_1, Y_1, X_2, Y_2, where the subscript refers to twin member. Below, "phenotypic covariance" refers to the covariance between variables within individuals (e.g. X_1-Y_1, X_2-Y_2), and "twin covariance" refers to the covariance within the same variable between two members of a twin-pair (X_1-X_2, Y_1-Y_2). The true covariance matrices are reported below. The population covariance matrix with the low phenotypic correlation is reported below the diagonal, and the population covariance matrix with the high phenotypic correlation is reported above the diagonal.

Population covariance matrix in MZ/DZ twins

	X1	Y1	X2	Y2
X1	1	.8/.8	.7/.3	.5/.2
Y1	.4/.4	1	.5/.2	.7/.3
X2	.7/.3	.3/.15	1	.8/.8
Y2	.3/.15	.7/.3	.4/.4	1

We used R library Norm to estimate the unconstrained MZ and DZ twin phenotypic covariance matrices. Norm optimizes the raw data likelihood function by means of the EM algorithm (Schafer, 1996).

Simulation study 2: Statistical power to detect genetic and environmental influences in the presence of missingness

Estimation of genetic and environmental variance components

The influence of the relative contributions of genetic and environmental factors to individual differences in ADHD can be inferred from the different level of genetic relatedness of MZ and DZ twins (Neale & Cardon, 1992). The variance may be due to additive genetic effects (A), dominant genetic effects (D) or shared environmental effects (C), and non-shared environmental (E) effects. The genetic effects (A and D) correlate 1 in MZ twins as they are genetically identical. In DZ twins, A correlates .5, because DZ twins share on average half of their segregating genes. The effects due to dominance correlate .25 in DZ twins. C correlates 1 in both MZ and DZ twins. E or non-shared environmental effects are, by definition, uncorrelated. All uncorrelated measurement error, if present, is also absorbed in the E term. Note that estimating C and D at the same time is not possible in a design using only data from MZ and DZ twins reared together. The decomposition of the phenotypic covariance matrix was based on Cholesky decompositions as shown in Figure 1. The variation in phenotype X (i.e., the variable on which the selection was based), and phenotype Y (i.e., the variable in which missingness is introduced) is influenced by A, C or D, and E. Simulation was based on two aetiological models: the ADE model and the ACE model. In the ADE model, variation in the phenotypes X and Y was for 50% explained by A, for 20% by D, and for 30% by E. The covariation between X and Y was for 71% explained by A and for 29% by D. The additive and dominant genetic correlations between X and Y were .80, and the non-shared environmental correlation of E was 0.

In the ACE model, A, C, and E explained 50%, 20% and 30%, respectively of the pehnotypic variance in X and Y. A and C explained 71% and 29%, respectively, of the phenotypic covariance between X and Y. The additive and shared environmental correlations between X and Y were .80, and the non-shared environmental correlation of E was 0. The ADE (below diagonal) and ACE (above diagonal) models imply the following covariance matrices in MZ and DZ pairs:

Population covariance matrix (MZ/DZ twins)

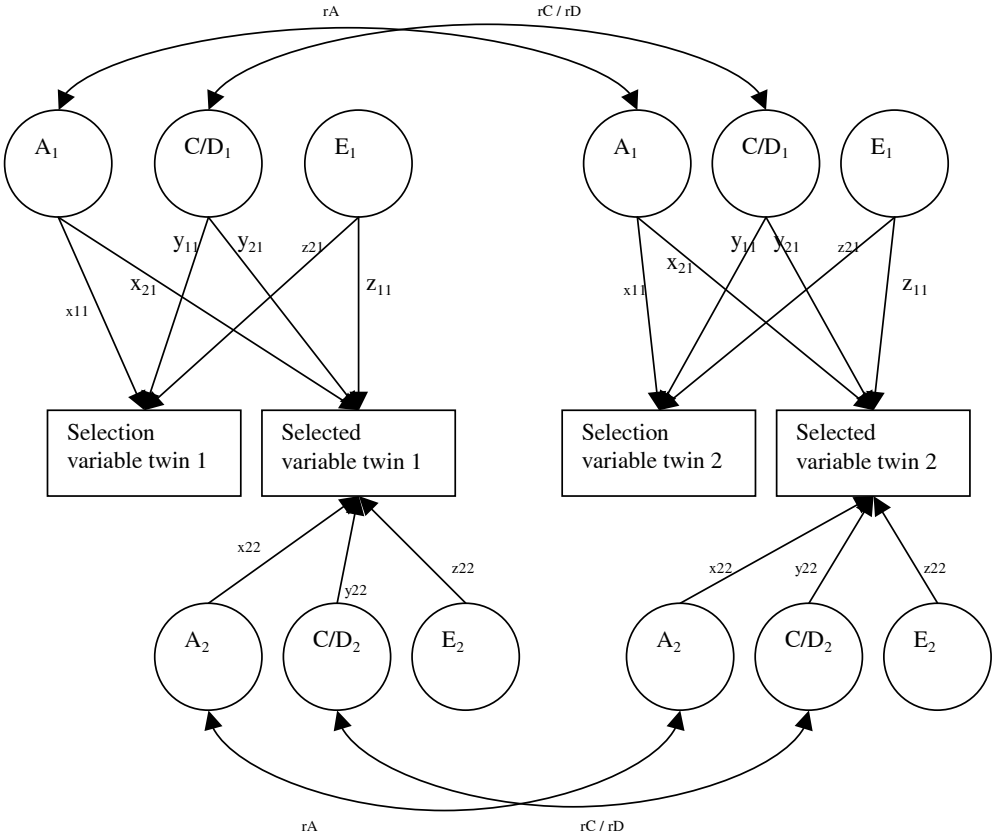
	X1	Y1	X2	Y2
X1	1	.56/.56	.70/.45	.56/.36
Y1	.56/.56	1	.56/.36	.70/.45
X2	.70/.30	.56/.24	1	.56/.56
Y2	.56/.24	.70/.30	.56/.56	1

Statistical analyses

The genetic model fitting was carried out in Mx (Neale et al., 2003). In the case of the ADE model, parameter estimates were obtained by fitting a multivariate ADE model and the statistical power to detect D on the selected variable was obtained by fixing the factor loadings y_{21} and y_{22} at zero (see Figure 1). In the ACE model, parameter estimates were obtained by fitting a multivariate ACE model and the statistical power to detect A and C was obtained by fixing the factor loadings x_{21} and x_{22}, and the loadings y_{21} and y_{22}, respectively.

Because of the implicit constraints on the parameters in the Cholesky decomposition, the null distribution of the likelihood ratio test for these parameters is not the expected central χ^2 (Carey, 2004; Dominicus et al., 2006b). Rather it is a mixture of χ^2 distributions (see Stram & Lee, 1994). To obtain some insight into the nature of this mixture, we performed a small simulation (detailed results are available on request). The results suggested strongly that the null-distribution is asymptotically a $\chi^2_{(1)}$-$\chi^2_{(2)}$ mixture with mixing proportions approximately equal to 50:50. These results tally with those of Stram and Lee (1994). The critical values associated with this mixture was estimated at 5.138 (=0.05) using a R program which is available on request from the corresponding author.

Figure 1. Multivariate twin model



Note: A=Additive genetic influences; C=Shared environmental influences; D=Dominant genetic influences; E=Non-shared environmental influences; rA=1 (.5) in MZ (DZ) twins; rD=1(.25) in MZ (DZ) twins; rC=1 in MZ (DZ) twins

RESULTS

Simulation study 1

In the first series of simulations we explored the influence of the size of the correlation between X and Y. The estimated variances and twin covariances in the variable with missingness (phenotype Y) are summarized in Tables 1a and 1b. The tables include information on the true variances and covariances, the thresholds that are used for selection and the percentage of missingness in MZ and DZ twins.

Except in the MCAR design, the larger SE's are associated with a lower correlation between phenotypes X and Y. Obviously, when data are MCAR, the phenotypic correlation between X and Y has no effect, because the selection is not based on X. The results further show that the parameter estimates are unbiased, and that the standard errors (SE; i.e., the standard deviation of the estimates) are comparable in all selection designs except the extreme discordant (ED) design. In this design, only a small number of MZ twins are selected (missingness is 93-94%). As a result, the variance and covariance estimates in MZ twins are biased and show large SE's. The lowest SE's are obtained in the individual selection design.

Table 1a. Estimation of the variance and covariance after the introduction of missingness: high correlation between the selection (X) and selected (Y) variable

	Thresh low/high	%mis MZ	%mis DZ	Var Y_1MZ	Var Y_1DZ	Covar X_1Y_2MZ	Covar X_1Y_2DZ	Covar Y_1Y_2MZ	Covar Y_1Y_2DZ
True value	-	-	-	1	1	.50	.20	.7	.3
Complete sample	-	0%	0%	1.00 (.03)	1.00 (.02)	.50 (.03)	.20 (.02)	.70 (.03)	.30 (.02)
Concordant low + individual high	-.5*/1.7	88%	88%	1.00 (.06)	1.00 (.04)	.50 (.03)	.20 (.02)	.70 (.06)	.30 (.03)
Extreme concordant	-1.05/1.05	85%	92%	1.00 (.06)	1.00 (.07)	.50 (.03)	.20 (.04)	.70 (.05)	.29 (.04)
Extreme discordant and concordant	-1.1/1.1	86%	92%	1.00 (.06)	1.00 (.04)	.50 (.03)	.20 (.02)	.70 (.04)	.30 (.03)
Extreme discordant	-.35/.35	93%	82%	1.05 (.33)	1.00 (.06)	.50 (.22)	.21 (.04)	.75 (.33)	.31 (.05)
Individual selection	-1.8/1.8	88%	87%	1.00 (.06)	1.00 (.03)	.50 (.03)	.20 (.02)	.70 (.05)	.30 (.02)
Missing completely at random	-/-	88%	88%	1.00 (.07)	.99 (.05)	.50 (.04)	.20 (.03)	.69 (.06)	.29 (.04)

* 25% of the pairs in which both members score below the threshold are randomly selected
Thresh low=lower threshold; thresh high=higher threshold; %mis=percentage of missingness; MZ=monozygotic; DZ=dizygotic.
A description of the selection designs is provided in the text. The variance estimates are based on the variance in Twin 1. In covar X_1Y_2 , the subscript refers to the twin member.

Table 1b. Estimation of the variance and covariance after the introduction of missingness: low correlation between the selection (X) and selected (Y) variable

	Thresh low/high	%mis MZ	%mis DZ	Var Y_1MZ	Var Y_1DZ	Covar X_1Y_2MZ	Covar X_1Y_2DZ	Covar Y_1Y_2MZ	Covar Y_1Y_2DZ
True value	-	-	-	1	1	.3	.15	.7	.3
Complete sample	-	0%	0%	1.00 (.03)	1.00 (.02)	.30 (.02)	.15 (.02)	.70 (.03)	.30 (.02)
Concordant low + individual high	-.5*/1.7	88%	88%	.99 (.08)	1.00 (.05)	.30 (.04)	.15 (.03)	.70 (.07)	.30 (.04)
Extreme concordant	-1.05/1.05	85%	92%	1.00 (.07)	1.00 (.06)	.30 (.04)	.16 (.06)	.70 (.06)	.30 (.06)
Extreme discordant and concordant	-1.1/1.1	86%	92%	1.00 (.08)	1.00 (.08)	.30 (.04)	.14 (.03)	.70 (.07)	.30 (.05)
Extreme discordant	-.35/.35	94%	82%	1.06 (.29)	1.00 (.07)	.29 (.29)	.15 (.07)	.75 (.28)	.30 (.05)
Individual selection	-1.8/1.8	88%	86%	1.00 (.07)	1.00 (.06)	.30 (.03)	.15 (.02)	.70 (.07)	.30 (.04)
Missing completely at random	-/-	88%	88%	1.01 (.08)	.99 (.06)	.30 (.06)	.15 (.04)	.71 (.06)	.30 (.05)

For explanation of table see note table 1a

Simulation study 2

As expected, given the FIML theory, the estimates of the genetic and environmental influences on phenotypes X, Y, and on the covariance between X and Y closely resemble the simulated values in the true ADE and ACE model (Tables 2a-b). The SE's of the standardized influences on phenotype Y are greater than the SE's of the standardized influences on phenotype X, as is to be expected, because missingness was limited to Y. Because MZ and DZ

variances are constrained to be equal, the SE's in the ED design are not much higher than the SE's in the other selection designs. The lowest SE's are again found in the individual selection design.

Table 2a. Standardized parameter estimates in ADE model

		Complete sample	Concordant low + individual high	Extreme Concordant	EDAC	Extreme Discordant	Individual selection	MCAR
Standardized variance component	True value	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
A variance X	.50	.49 (.05)	.50 (.06)	.50 (.05)	.50 (.06)	.50 (.05)	.50 (.06)	.50 (.06)
A covariance XY	.71	.72 (.07)	.74 (.12)	.72 (.19)	.73 (.13)	.70 (.18)	.74 (.10)	.72 (.15)
A variance Y	.50	.50 (.06)	.50 (.12)	.48 (.16)	.48 (.13)	.47 (.16)	.50 (.11)	.47 (.13)
D variance X	.20	.21 (.05)	.20 (.06)	.20 (.06)	.20 (.06)	.20 (.05)	.20 (.06)	.20 (.06)
D covariance XY	.29	.29 (.08)	.26 (.12)	.27 (.18)	.26 (.13)	.30 (.18)	.27 (.11)	.28 (.16)
D variance Y	.20	.20 (.06)	.20 (.12)	.22 (.17)	.22 (.13)	.23 (.16)	.21 (.11)	.22 (.13)
E variance X	.30	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)
E covariance XY	.00	.00 (.01)	.00 (.03)	.01 (.04)	.01 (.03)	.00 (.03)	.00 (.03)	.01 (.03)
E variance Y	.30	.30 (.01)	.30 (.03)	.30 (.03)	.30 (.02)	.29 (.03)	.30 (.03)	.30 (.03)

A=additive genetic effects, D=dominant genetic effects; E=non-shared environmental effects; EDAC=extreme discordant and concordant; MCAR=missing completely at random

Table 2b. Standardized parameter estimates in ACE model

		Complete sample	Concordant low + individual high	Extreme Concordant	EDAC	Extreme Discordant	Individual selection	MCAR
Standardized variance component	True value	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
A variance X	.50	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)	.50 (.03)
A covariance XY	.71	.71 (.04)	.73 (.08)	.72 (.11)	.72 (.09)	.73 (.10)	.71 (.06)	.72 (.08)
A variance Y	.50	.50 (.03)	.49 (.08)	.49 (.09)	.50 (.08)	.49 (.10)	.50 (.07)	.50 (.08)
C variance X	.20	.20 (.02)	.20 (.02)	.20 (.03)	.20 (.03)	.20 (.03)	.20 (.02)	.20 (.03)
C covariance XY	.29	.29 (.03)	.27 (.06)	.28 (.10)	.28 (.07)	.27 (.10)	.29 (.05)	.29 (.07)
C variance Y	.20	.20 (.02)	.20 (.06)	.21 (.08)	.20 (.07)	.21 (.10)	.20 (.06)	.20 (.07)
E variance X	.30	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)	.30 (.01)
E covariance XY	.00	.00 (.01)	.00 (.03)	.00 (.04)	-.01 (.04)	-.01 (.03)	.00 (.02)	.00 (.03)
E variance Y	.30	.30 (.01)	.30 (.03)	.30 (.03)	.30 (.02)	.30 (.04)	.30 (.03)	.30 (.03)

A=additive genetic effects, C=shared environmental effects; E=non-shared environmental effects; EDAC=extreme discordant and concordant; MCAR=missing completely at random

Table 3 shows the statistical power to detect the two paths of A, C, and D loading on phenotype Y. The decrease in statistical power as a result of the introduction of missingness can be derived from a comparison of the mean χ^2 differences with the complete sample (i.e., the value when no missingness is introduced). A lower mean difference in χ^2 is associated with a lower statistical power. In addition to the mean difference in χ^2, we included the proportion of simulations in which the null hypothesis (i.e., no effect of A, C, or D) was rejected.

Table 3. Effect of missingness on the statistical power to detect additive genetic effects (A), dominant genetic effects (D), and shared environmental effects (C)

	Detection of D (in ADE model)		Detection of C (in ACE model)		Detection of A (in ACE model)	
	Mean χ^2(SD)	Statistical power	Mean χ^2(SD)	Statistical power	Mean χ^2(SD)	Statistical power
Complete sample	17.89 (7.91)	.98	77.75 (16.72)	1.00	432.23 (36.89)	1.00
Concordant low + individual high	5.25 (3.81)	.43	20.30 (8.86)	.99	100.00 (18.44)	1.00
Extreme concordant	3.02 (2.82)	.16	10.33 (6.06)	.77	67.90 (17.58)	1.00
Extreme discordant and concordant	4.71 (3.59)	.36	13.98 (5.57)	.96	75.01 (17.51)	1.00
Extreme discordant	4.51 (4.11)	.33	8.77 (5.68)	.65	99.63 (17.59)	1.00
Individual selection	7.06 (4.82)	.60	29.46 (10.13)	1.00	119.63 (20.85)	1.00
Missing completely at random	4.34 (3.26)	.36	14.09 (6.92)	.90	77.43 (15.73)	1.00

Note: Statistical power is power with alpha=.05, df=2

The statistical power to detect genetic dominance is low in all six selection designs. In the case of complete data, the power to detect a D effect of 20% (variance explained) equals 0.98. After the introduction of 88% missingness, the statistical power drops to 0.16-0.60. The power is highest in the individual selection design and lowest in the EC design. In the complete sample, the power to detect a C component of 20% is 1. The power drops to 0.65-1.0 after the introduction of missingness. Finally, the power to detect an A component of 50% does not deviate from the theoretical value of 1.0 in the absence of missingness, irrespective of the selection design. Clearly the effect size of A is too large to pick up any differences in power.

DISCUSSION

In the current paper we focused on the situation, where a relatively cheap measure (X) is measured in a large representative sample, and a more expensive measure (Y) is measured in a subset of the total population. Little & Rubin (2002) have shown that, provided that the missing data are either MAR or MCAR, FIML produces an unbiased estimate of the population covariance matrix. The effect of planned missingness on the statistical power to detect genetic and environmental influences has yet to be studied. The goal of the paper was to investigate the statistical power in various selection designs, which give rise to planned missingness.

The effect of the correlation between the two measurements

We first investigated the effect of the correlation between the cheap measure (phenotype X) and the expensive measure (phenotype Y) on the SE's of the twin (co)variances for Y. If the phenotypic correlation between X and Y is zero, the selection of subjects results in data MCAR. Therefore, the selection on extreme scores on X would not increase the statistical power to detect twin covariances. As the phenotypic correlation between X and Y increases, the subjects who are selected based on extreme scores on X will also obtain more extreme scores on Y. Given the fact that a covariance can be more precisely estimated when extreme scores are sampled, we would expect to observe lower SE's of the estimated twin covariances (i.e., the covariances X_1Y_2 and Y_1Y_2) as the correlation between X and Y is increases. Inspection of the results of simulation study 1 show that lower SE's are indeed associated

with a higher phenotypic correlation between X and Y. In conclusion, it was shown that the selection based on phenotype X is most effective when the correlation between phenotypes X and Y is high.

The effect of missingness on the statistical power to detect genetic and environmental influences in various selection designs

In simulation study 1, we showed that missingness resulted in relatively minor increases in SE's of the variances and twin covariances in five out of six selection designs. The ED design was the only one that resulted in a large increase of the SE's in the MZ samples. The large SE's in MZ twins are the result of the small number of extremely discordant pairs in MZ twins compared to DZ twins. In simulation study 2, model specification ensured that the variances of MZ and DZ twins were equated, and consequently the SE's did not deviate much from the SE's in the other designs. Because of the more precise estimate of the variance, the covariance estimates also improved. Estimation of genetic and environmental influences resulted in SE's that were similar in the ED, EC, EDAC, and MCAR designs. The smallest SE's were obtained using the individual selection design. In the first design, in which pair-wise and individual selection was combined, the SE's were intermediate between the individual and the other designs.

In all instances, missingness resulted in decreased values of ^{2} compared to the values in the complete sample. However, the statistical power to detect additive genetic influences is high, irrespective of the selection design that was used. Because of the relatively large sample size of phenotype X (4000 DZ and 2000 MZ pairs), and the large effect size (50%), all selection designs provided excellent power. The power to detect shared environmental or dominant genetic influences, which both explain 20% of the variation, is much lower. The best selection design, with respect to the statistical power to detect these influences, is the individual selection design. However, even with this design, the power to detect genetic dominance is only 60%. The power to detect shared environmental influences is satisfactory (above 80%) in most designs.

The individual selection design is the best design with respect to the detection of genetic and environmental contributions to the covariance between phenotypes X and Y, and to the variance of Y.In the individual selection design, twins with a moderate score are included if their co-twin has an extreme score. The inclusion of moderate scores seems to improve the estimation of the regression of X_1 on Y_2 and of Y_1 on Y_2. To examine this hypothesis, we simulated data and estimated the variances and covariances in MZ and DZ twins as we did in simulation study 1. As in the other situations, 88% missingness was introduced. Of the total population, 5% of the subjects was randomly selected and 7% was selected in accordance with the EC design. As a result of the inclusion of a small random sample, the standard errors of the variances and covariances were lower both in MZ and DZ twins compared to both the EC design and the MCAR design. An explanation for this finding may be that the inclusion of subjects with moderate scores increases the variation in Y. Although the individual selection design was the best design in this study, most studies are not only performed with the goal of variance decomposition into latent genetic and environmental influences. An additional purpose may be the detection of effects of measured genetic polymorphisms. For this purpose, the ED and EDAC designs are more suitable (Gu et al., 1996). The current simulations show that the detection of shared environment may be problematic in the ED design, but not the EDAC design. In both designs, the power to detect genetic dominance is low. Because the statistical power to detect dominance is also quite low in the individual selection design, the EDAC design is a good choice both for traditional genetic covariance structure analysis, and for QTL analysis. For the detection of genetic dominance, larger sample sizes are needed.

The fact that the statistical power to detect A and C is satisfactory, even with a percentage of missingness of 88% (resulting in a sample size of only 720 pairs), is promising for longitudinal studies in which attrition may lead to missing phenotypic data in a subset of the sample. Usually, non-response rates will be much lower than 80%, which suggests that the power to detect genetic and environmental influences is good, given that the original (total) sample size is not very small. Considering the effect of missingness as a result of attrition, it should be kept in mind that we restricted ourselves to the discussion of response models under the assumption that the data are MAR or MCAR. When the data are actually missing not at random (MNAR), maximum likelihood estimation can not correct for the missingness, and genetic and environmental influences may either be overestimated or underestimated, depending on the nonresponse model (Taylor, 2004; Dominicus et al., 2006a).

Limitations

The results of the present study should be interpreted in consideration of the following limitations. First, we assumed that the data are either MAR or MCAR. This does not affect the results in the present studies, because the selection strategies that were applied guaranteed MAR. In other situations, however, the mechanism giving rise to missingness may be unknown which would result in data MNAR. As discussed above, in these instances, parameters may be biased, when data are actually MNAR. Second, we assumed that phenotypes X and Y are multivariate normally distributed. If the true distribution of X is not normal, for example as a result of censoring or truncation, the selection may not be optimal and the decrease in statistical power will be greater, than observed in the current analyses.

Conclusions

The decrease in statistical power to detect additive genetic and shared environmental influences as a result of 88% missingness is small, although the power to genetic dominance is poor after the introduction of missingness. Larger sample sizes (less extreme selection) would have to be considered if the presence of genetic dominance is expected. The best selection design with respect to the specific purpose of variance decomposition into latent genetic and environmental influences is the selection design in which twin-pairs are selected on an individual basis. However, because the statistical power under the EDAC design is also satisfactory, this design could be preferable, if the aim of the study was both to investigate the genetic covariance structure of X and Y, and to find QTLs contributing to the individual differences in Y.

REFERENCE LIST

Carey, G. (2004). Cholesky problems. *Behavior Genetics, 34,* 633.

Derks, E. M., Hudziak, J. J., Dolan, C. V., Ferdinand, R. F., & Boomsma, D. I. (2006). The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. *Comprehensive Psychiatry, 47,* 116-122.

Dolan, C., Sluis van der, S., & Grasman, R. (2005). A note on normal theory power calculation in SEM with data missing completely at random. *Structural Equation Modeling-A Multidisciplinary Journal, 12,* 245-262.

Dominicus, A., Palmgren, J., & Pedersen, N. L. (2006a). Bias in variance components due to nonresponse in twin studies. *Twin Research and Human Genetics, 9,* 185-193.

Dominicus, A., Skrondal, A., Gjessing, H. K., Pedersen, N. L., & Palmgren, J. (2006b). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics, 36,* 331-340.

Eaves, L. & Meyer, J. (1994). Locating Human Quantitative Trait Loci - Guidelines for the Selection of Sibling Pairs for Genotyping. *Behavior Genetics, 24,* 443-455.

Gu, C., Todorov, A., & Rao, D. C. (1996). Combining extremely concordant sibpairs with extremely discordant sibpairs provides a cost effective way to linkage analysis of quantitative trait loci. *Genetic Epidemiology, 13,* 513-533.

Heath, A. C., Madden, P. A. F., & Martin, N. G. (1998). Assessing the effects of cooperation bias and attrition in behavioral genetic research using data-weighting. *Behavior Genetics, 28,* 415-427.
Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data.* (2 ed.) New York: Wiley and Sons.
Martin, N. G. & Eaves, L. J. (1977). Genetic-Analysis of Covariance Structure/e. *Heredity, 38,* 79-95.
Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA: Department of Psychiatry.
Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publisher.
Risch, N. & Zhang, H. P. (1995). Extreme Discordant Sib Pairs for Mapping Quantitative Trait Loci in Humans. *Science, 268,* 1584-1589.
Rubin, D. B. (1976). Inference and Missing Data. *Biometrika, 63,* 581-590.
Schafer, J. L. (1996). *Analysis of Incomplete Multivariate Data by Simulation*. New York: Chapman and Hall.
Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147-177.
Stram, D. O. & Lee, J. W. (1994). Variance-Components Testing in the Longitudinal Mixed Effects Model. *Biometrics, 50,* 1171-1177.
Taylor, A. (2004). The consequences of selective participation on behavioral-genetic findings: Evidence from simulated and real data. *Twin Research, 7,* 485-504.
Venables, W. N., Smith, D. M., & R Development Core Team (2002). *An introduction to R*. Bristol, United Kingdom: Network Theory Limited.

CHAPTER

7

EFFECTS OF CENSORING ON PARAMETER ESTIMATES AND POWER IN GENETIC MODELING

This paper is published as Derks E.M., Dolan C.V., and Boomsma D.I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research, 7,* 659-669.

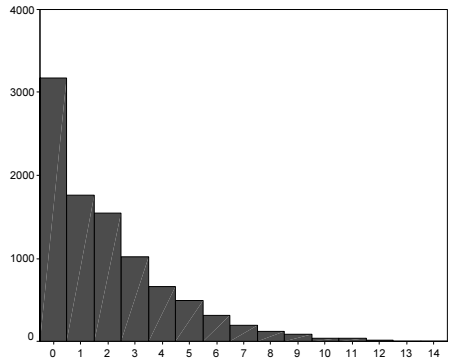
ABSTRACT

Genetic and environmental influences on variance in phenotypic traits may be estimated with normal theory Maximum Likelihood (ML). However, when the assumption of multivariate normality is not met, this method may result in biased parameter estimates and incorrect likelihood ratio tests. We simulated multivariate normal distributed twin-data under the assumption of three different genetic models. Genetic model fitting was performed in six data sets: multivariate normal data, discrete uncensored data, censored data, square root transformed censored data, normal scores of censored data, and categorical data. Estimates were obtained with normal theory ML (data sets 1-5) and with categorical data analysis (data set 6). Statistical power was examined by fitting reduced models to the data. When fitting an ACE model to censored data, an unbiased estimate of the additive genetic effect was obtained. However, the common environmental effect was underestimated and the unique environmental effect was overestimated. Transformations did not remove this bias. When fitting an ADE model, the additive genetic effect was underestimated while the dominant and unique environmental effects were overestimated. In all models, the correct parameter estimates were recovered with categorical data analysis. However, with categorical data analysis, the statistical power decreased. The analysis of L-shaped distributed data with normal theory ML results in biased parameter estimates. Unbiased parameter estimates are obtained with categorical data analysis, but the power decreases.

INTRODUCTION

Thanks to computational and methodological advances over the last few decades, genetic covariance structure modeling in genetically informative samples is relatively straightforward. Estimates of genetic and environmental variance components may be obtained readily using programs like Mx (Neale et al., 2003), Lisrel (Jöreskog & Sörbom, 1996a), or Mplus (Muthén & Muthén, 2001; Prescott, 2004). The dominant method of estimation is normal theory Maximum Likelihood (normal theory ML), which is based on the assumption of multivariate normality. Unfortunately, the distribution of phenotypic data may display a large degree of skewness and kurtosis, which renders the choice of normal theory ML to estimate parameters suboptimal. The problem of non-normality is acute in the study of symptom data, where the distribution of observed symptoms is often L-shaped, due to the fact that the vast majority of subjects displays a few or no symptoms (Oord van den et al., 2003). Failure to account for non-normality, may lead to biased parameter estimates and incorrect likelihood ratio tests (Amos, 1994). There are many possible causational factors for the presence of non-normality. These can be divided into two categories: a) factors that lead to a non-normal distribution of the latent trait; and b) factors that lead to a non-normal distribution of the measured indicators of a normally distributed latent trait. If the distribution of the latent trait is not normal, a possible solution is to adopt a more appropriate distribution (e.g., Poisson). If the latent trait is normally distributed, but the observed trait is not, for example due to censoring, a possible solution is to correct the observed data for the censoring event.Oord van den et al. (2003) proposed that the latent distribution of L-shaped behavioral checklist data is normal. They examined this hypothesis by means of Item Response Theory (IRT, Hambleton & Swaminathan, 1985) and found that a model that allowed for non-normality in the latent distribution did not provide a better fit than a model that did not allow for non-normality. In other words, they found no evidence against a normal latent distribution. Therefore, we assume that the L-shaped distribution of behavioral checklist data belongs to the second category. This seems plausible because questions in behavior checklists are often developed with the purpose of determining the degree of behavioral dysfunctioning. In the latent normal distribution, children with well adapted behavior may be found at the left tail of the distribution, and children with dysfunctional behavior may be found at the right tail of the distribution. Because of the focus on behavioral dysfunction, variation in the right tail in the distribution is measured while variation in the left tail of the distribution is not. This results in a L-shaped observed distribution. An example of such a distribution is shown in Figure 1. This figure illustrates the degree of sleep problems in three-year-old children. The distribution clearly is not normally distributed, probably caused by censoring.

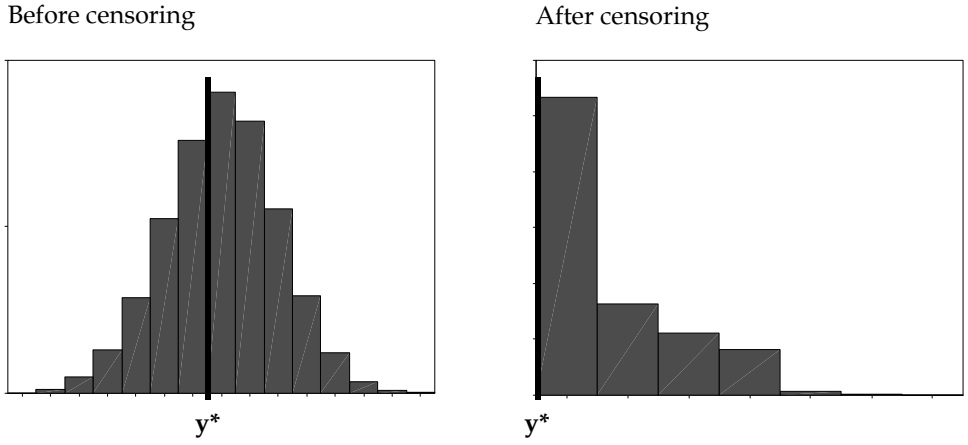
Figure 1. Distribution of Maternal Reports of Sleep Problems



Note: Graph is based on 9,415 first-born twins; data of second born twins are similar

Censored data arise if values below (or above) a certain threshold y* are observed at y*. As a result, below (or above) this threshold, variation in the distribution of the latent trait is unobserved, and the observed distribution is skewed. The effect of censoring from below is illustrated in Figure 2.

Figure 2. A Graphical Representation of the Censoring Process



Note: Before censoring, the complete normal distribution is shown. After censoring, values below y* are observed at value y*

There are numerous practical examples of censored distributions, in many different fields of inquiry (e.g., economics, medicine, and the social sciences). One of the earliest attempts to address censoring is that of (Tobin, 1958). He studied the demand for various categories of capital goods such as automobiles. Many households report zero expenditures in a given year. Among the households that made an expenditure, there is large variability in amount. The observed demand for capital goods in a given year is therefore censored below. An example from the field of medicine, is the assessment of coronary artery calcification (Epstein et al., 2003). Coronary artery calcification is only assumed to be present when it exceeds a certain threshold. Below this threshold, the level is assumed to be zero (Bielak et al., 2001). Finally, censoring is present in behavioral ratings (Nagin & Tremblay, 1999; Rietveld et al., 2003).

There are several methods to correct for non-normality. First, the data may be transformed in order to achieve normality. Often applied transformations are the logarithmic and square root transformations (Lynch & Walsh, 1998). Transformations may be substantively motivated, e.g., the use of a logarithmic transformation when a trait is measured on a geometric scale instead of an arithmetic scale (Falconer, 1965). Examples of such traits are body weight and growth. Alternatively, transformations may be *ad hoc*. An example of this is the single parameter Box-Cox transformation, where a parameter is optimized to achieve normality. In the case of behavioral rating data, the transformation is generally *ad hoc*. The calculation of so-called normal scores (Jöreskog et al., 1999) may also be viewed as an *ad hoc* transformation. The aim of this transformation, which is based on the assumption of an underlying normal distribution, is to render the skewness and kurtosis of the data consistent with a normal distribution (i.e., values of approximately 0 and 3, respectively). Unfortunately, given major deviations from normality, transformations may fail to achieve this aim.

Second, we may assume that the observed data is measured on an ordinal scale instead of a continuous scale and adopt a method of estimation which is suitable for ordinal data. Parameter estimates of the genetic model fitting can be obtained under the assumption of an underlying continuous liability distribution that has one or more thresholds that define categories. This technique was independently developed by Crittenden (1961) and Falconer (1965) (Lynch and Walsh, 1998, pp 730). The estimates obtained with ordinal data analysis should be unbiased, but the analyses may be computationally more demanding, especially when the number of categories is large. Another disadvantage is the potential presence of empty cells in the contingency tables. For example, the contingency table of a highly heritable trait in MZ twins is likely to have some empty off-diagonal cells.

The purpose of this study is to examine the effect of censoring on the results of genetic modeling. We assume that a latent trait is normally distributed, and that censoring arises due to failure of the measurement instrument to detect values smaller than some general threshold y* (see Figure 2). Three methods which may be used to deal with non-normal data are compared in a simulation study. Two of these methods concern *ad hoc* transformations: a square root transformation and the computation of normal scores. The third method is the analysis of categorical data which is based on the liability threshold model. Finally, we apply these methods to real-life data on sleep problems in a large sample of 3-year-old twins.

METHODS

GENETIC MODELING

Variation in a phenotypic trait can be decomposed into latent genetic and environmental components. The decomposition of variance may be achieved by analysing data of pairs of individuals who differ in their degree of genetic relatedness. The twin design is a well known example of this approach. Monozygotic (MZ) twins are genetically identical, while dizygotic (DZ) twins on average share half of their segregating genes. Limiting the genetic decomposition of phenotypic variance to additive genetic (A) effects and dominant genetic (D) effects, the fact that MZ twins are genetically identical implies that they share all the additive and dominant genetic variance. DZ twins on average share half of the additive genetic and one quarter of the dominant genetic variance (Lynch & Walsh, 1998). The environmental decomposition of the phenotypic variance is into shared environmental variance and non-shared, or specific, environmental variance. The environmental effects shared by two members of a twin pair (C) are by definition perfectly correlated in both MZ and DZ twins. The non-shared environmental effects (E) are by definition uncorrelated between twin pair members. Estimates of the non-shared environmental variance usually include

measurement error (Plomin et al., 2001; Neale & Cardon, 1992). In fitting models to twin data, it is not possible to estimate the effects of all components of variance (Va, Vd, Vc, and Ve) simultaneously. Specifically, one cannot estimate Vd and Vc simultaneously due to reasons of identification.

Simulation study

Data were simulated in accordance with three models. In Models 1 and 2, the covariance structure of MZ and DZ twins was attributable to A, C, and E. The values of Va, Vc, and Ve in Model 1 equalled .50, .20, and .30, respectively. In Model 2, the values of Va, Vc, and Ve equalled .20, .50, and .30. In model 3, the covariances were influenced by A, D, and E. The values of the variance components Va, Vd, and Ve equalled .45, .25, and .30.

The population covariance matrices of MZ and DZ twins can be calculated under assumption of these three theoretical models. We assumed that there is no assortative mating, epistasis, gene-environment interaction or gene-environment correlation (Lynch & Walsh, 1998). Under these assumptions, the covariances of MZ twins are Va + Vc + Vd, and the covariances of DZ twins are .5*Va + Vc + .25*Vd. The variances equal Va + Vc + Vd + Ve for both MZ and DZ twins. In the simulation study, the covariances of MZ and DZ twins were .70 and .45 (Model 1), .70 and .60 (Model 2), and .70 and .2875 (Model 3), respectively. All variances equalled 1.

In most European countries, the number of DZ twins is larger than the number of MZ twins. For example, in the Netherlands Twin Register, the number of DZ twins is about twice the number of MZ twins. Therefore, the number of DZ twins in the simulation study was also twice the number of MZ twins. We simulated data of 3,000 MZ twin pairs and 6,000 DZ twin-pairs. This sample size is representative for the sample size of twin registers such as the Netherlands Twin Register (Boomsma, 1998; Boomsma et al., 2002). The simulation study comprised 1,000 replicates.

Data simulation was performed in R (Venables et al., 2002). Six different data sets were generated. The distributions of these data sets are shown in Figure 3a to Figure 3f. First, multivariate standard normal distributed data (Figure 3a) were simulated with the MASS package which is available in the R library. Second, because observational data are usually discrete, the multivariate normal data were discretized (Figure 3b). The number of categories was 15. The values of the categories were chosen arbitrarily: a value of -7 was assigned to the lowest category, and a value of +7 was assigned to the highest category.

Third, the discrete data were censored (Figure 3c). The value of the censor was equal to 0. The values of all data points below 0, which made up 39% of the total data set, were therefore reassigned to 0 in the censored data set. This percentage was chosen because the resulting distribution resembled the distribution of behavioral checklist data in terms of skewness and kurtosis. The fourth, and fifth data sets were created by applying two transformations to the censored data. The first transformation was a square root transformation (Figure 3d). The second transformation was the computation of normal scores (Figure 3e). This transformation renders the skewness and kurtosis of the data as close as possible to 0 and 3, the expected values of the skewness and kurtosis when the distribution is normal. The computation of normal scores is implemented in Prelis (Jöreskog & Sörbom, 1996b). For this simulation study, the Prelis procedure was implemented in R. The R syntax is available on request. Finally, the number of categories of the censored data was decreased to four (Figure 3f). In contrary to the five other data sets, where we applied normal theory ML, these data were treated as categorical data. All analyses were performed on raw data.

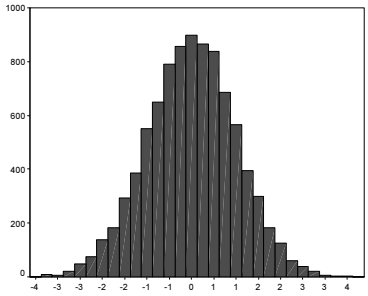
GENETIC ANALYSES OF SLEEP PROBLEMS

Subjects

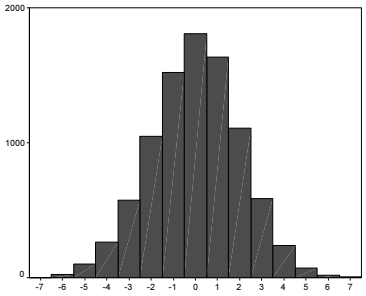
The subjects were all registered at birth with the Netherlands Twin Registry (Boomsma et al., 2002). In the present study, we have assessed a sample of Dutch twin pairs whose mothers reported on their sleep problems when the twins were three years old. These twins were all born between 1986 and 1997. The sample that was used for the genetic analyses consisted of 6,375 MZ twins and 12,192 DZ twins. Zygosity diagnosis was assessed with the use of a 10-item questionnaire. This procedure allows an accurate determination of zygosity of nearly 95%. It is described in more detail in Rietveld et al. (2000). For a more detailed description of the sample, see Derks et al. (2004).

Figure 3. Distributions of Six Simulated Data Sets. Data sets were Generated from a Bivariate Normal Distribution (a) and then Discretized (b). Next, Data were Censored (c). These Censored Data were either Transformed by Square Root Transformation (d) or to Normal Scores (e). Finally, the Number of Categories of the Censored Scores was Limited to Four (f)

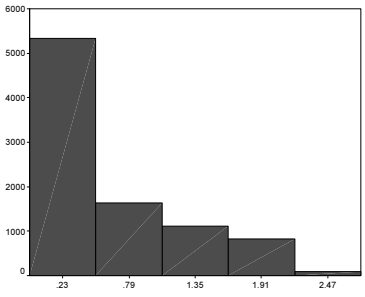
a. Normal distributed scores



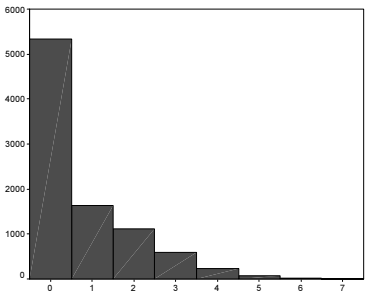
b. Discretized normal distributed scores



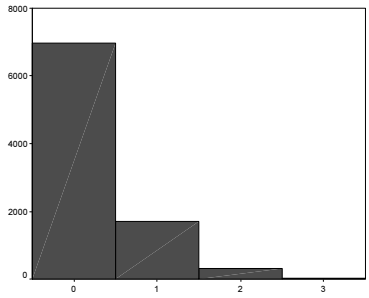
c. Censored scores



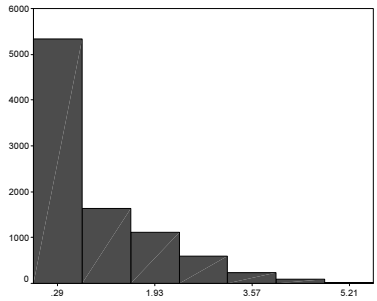
d. Square root transformed censored scores



e. Normal scores of censored data



f. Categorical scores of censored data



Measure

The Child Behavior Checklist (CBCL/2-3) is a standardized questionnaire for parents to report the frequency and intensity of behavioral and emotional problems exhibited by their child in the past six months (Achenbach, 1992). It contains 100 items that measure problem behavior; the items are rated on a 3-point scale ranging from "not true", "somewhat or sometimes true" to "very true or often true". The CBCL measures the number of symptoms on seven behavioral syndromes, including sleep problems (7 items) (Koot et al., 1997). The distribution of sleep problems is shown in Figure 1 (first-borns only to save space).

RESULTS

Simulation study

The descriptive statistics of the simulated data before and after transformation are reported in Table 1. These descriptives are reported for one replication and a single twin only. As expected, the skewness and kurtosis of the non-censored continuous data did not deviate significantly from the expected values of 0 and 3. After discretization, the variance increased as a result of the larger range of values but the skewness and kurtosis were unaffected. After censoring, the skewness and kurtosis were both positive and deviated significantly from 0 and 3. In addition, the mean of the data increased and the variance decreased in comparison to the non-censored discrete data. Both the square root transformed data and the normal scores showed less skewness and kurtosis than the untransformed censored data, but their values still deviated significantly from 0 and 3.

Table 1. Descriptive Statistics of Simulated Data

Data set	Mean	Standard Deviation	Skewness	Kurtosis
1. Non-censored continuous data	.000	.996	-.043	2.985
2. Non-censored discrete data (15 categories)	.004	2.012	-.053	2.991
3. Censored discrete data (8 categories)	.790	1.172	1.603	5.293
4. Square root transformed censored data	1.639	.317	1.277	3.779
5. Normal scores of censored data	.786	1.172	.946	2.794
6. Categorical data (4 categories based on censored data)	Threshold 1=0.755 Threshold 2=1.787 Threshold 3=2.748			

Data Sets were Generated from a Bivariate Normal Distribution (1) and then Discretized (2). Next, Data were Censored (3). These Censored Data were either Transformed by Square Root Transformation (4) or to Normal Scores (5). For the 6th Data Set (Four Cate-

gories based on Censored Data), the Thresholds are Given. The Number of Replications is 1,000, but Descriptives are Given for a Single Replication and for a Single Twin only. The Number of Twins is 9,000.

THE PARAMETER ESTIMATES OF THE ACE MODELS

Table 2a and Table 2b show the mean point estimates of the standardized variance components and their standard deviations in the 1,000 replications. The mean point estimate of the categorical data analyses was based on slightly fewer than 1,000 replications, because the minimalization of the likelihood failed in one of the replications.

As expected, the analysis of non-censored continuous data produced the correct mean parameter estimates in both models. Discretization did not affect the mean or standard deviation of the standardized parameter estimates. After censoring, the estimate of Va was unbiased but Vc was underestimated and Ve was overestimated. A square root transformation or a transformation to normal scores did not improve the parameter estimates. In contrast, when the categorical data were analysed using the threshold model, the correct parameter estimates were recovered. However, as is to be expected given the reduced amount of information, the standard errors of the parameter estimates increased which resulted in wider confidence intervals and less precise estimates.

THE PARAMETER ESTIMATES OF THE ADE MODEL

The results of the ADE model (Table 2c) are in agreement with the results of the ACE models. The non-censored continuous data and the non-censored discrete data both recovered the correct parameter estimates. The analyses of the censored untransformed data, the square root transformed data, and the normal scores lead to biased parameter estimates. Va was underestimated, while Vd and Ve were both overestimated. When the categorical data option in Mx was used, the unbiased parameter estimates were obtained, but again with increased standard errors. The mean point estimate of the categorical data analyses was again based on slightly less than 1,000 replications, because the minimalization of the likelihood failed in six of the replications.

The underestimation of Va in the ADE model was large compared to the other deviations. While the underestimation of Vc in the ACE model and the overestimation of Vd and Ve in the ADE model varied between 5-10% of the variance, the underestimation of Va in the ADE model was about 20%.

Table 2a. Estimates of Standardized Genetic and Environmental Influences in the Six Simulated Data Sets, Averaged over 1,000 replications.
True values of Va, Vc, and Ve are .50, .20, and .30 respectively.

DATA SET	METHOD OF ANALYSIS	MEAN STAND. VA (SD)	MEAN STAND. VC (SD)	MEAN STAND. VE (SD)
1. Non-censored continuous data	Normal theory ML	.501 (.025)	.199 (.021)	.300 (.008)
2. Non-censored discrete data (15 categories)	Normal theory ML	.491 (.025)	.195 (.021)	.314 (.008)
3. Censored discrete data (8 categories)	Normal theory ML	.496 (.040)	.116 (.031)	.389 (.014)
4. Square root transformed censored data	Normal theory ML	.488 (.037)	.122 (.029)	.390 (.014)
5. Normal scores of censored data	Normal theory ML	.471 (.036)	.128 (.029)	.401 (.013)
6. Categorical data (4 categories based on censored data) [a]	Categorical data analysis	.501 (.053)	.200 (.042)	.300 (.019)

Note:Stand.=standardized; Va=variance explained by additive genetic effects; Vc=variance explained by common environmental effects; Ve=variance explained by unique environmental effects

[a] The results of the categorical data analyses are based on 999 replications due to minimalization problems in 1 replication

Table 2b. Estimates of Standardized Genetic and Environmental Influences in the Six Simulated Data Sets, Averaged over 1,000 Replications.
True values of Va, Vc, and Ve are .20, .50, and .30 respectively

DATA SET	METHOD OF ANALYSIS	MEAN STAND. VA (SD)	MEAN STAND. VC (SD)	MEAN STAND. VE (SD)
1. Non-censored continuous data	Normal theory ML	.200 (.020)	.500 (.016)	.300 (.008)
2. Non-censored discrete data (15 categories)	Normal theory ML	.196 (.021)	.490 (.017)	.314 (.009)
3. Censored discrete data (8 categories)	Normal theory ML	.205 (.036)	.406 (.027)	.389 (.015)
4. Square root transformed censored data	Normal theory ML	.203 (.034)	.407 (.025)	.390 (.014)
5. Normal scores of censored data	Normal theory ML	.194 (.034)	.404 (.025)	.402 (.014)
6. Categorical data (4 categories based on censored data) [a]	Categorical data analysis	.199 (.048)	.501 (.036)	.300 (.018)

Note:Stand.=standardized; Va=variance explained by additive genetic effects; Vc=variance explained by common environmental effects; Ve=variance explained by unique environmental effects

[a] The results of the categorical data analyses are based on 999 replications due to minimalization problems in 1 replication

POWER ANALYSES

One of the desirable features of an estimation method is that it should produce unbiased parameter estimates. Another important feature is statistical power. In this section, we compare the power of the different methods. To this end, we compared the fit of the true ACE model to the fit of an AE model and the fit of a CE model. We also compared the fit of the ADE model to the fit of an AE model. We did not compare the fit of the ADE model to the fit of a DE model because the presence of dominant genetic influences in the absence of additive genetic influences is biologically implausible (Falconer & Mackay, 1996).

Table 2c. Estimates of Relative Genetic and Environmental Influences in the Six Simulated Data Sets, Averaged over 1,000 Replications. True values of Va, Vd, and Ve are .45, .25, and .30 respectively

DATA SET	METHOD OF ANALYSIS	MEAN STAND. VA (SD)	MEAN STAND. VC (SD)	MEAN STAND. VE (SD)
1. Non-censored continuous data	Normal theory ML	.448 (.047)	.252 (.049)	.300 (.008)
2. Non-censored discrete data (15 categories)	Normal theory ML	.438 (.047)	.248 (.049)	.248 (.008)
3. Censored discrete data (8 categories)	Normal theory ML	.273 (.059)	.338 (.063)	.389 (.014)
4. Square root transformed censored data	Normal theory ML	.286 (.057)	.323 (.061)	.391 (.013)
5. Normal scores of censored data	Normal theory ML	.299 (.056)	.300 (.060)	.402 (.013)
6. Categorical data (4 categories based on censored data) [a]	Categorical data analysis	.446 (.087)	.254 (.093)	.300 (.018)

Note:Stand.=standardized; Va=variance explained by additive genetic effects; Vd=variance explained by dominant genetic effects; Ve=variance explained by unique environmental effects

[a] The results of the categorical data analyses are based on 994 replications due to minimalization problems in 6 replications

Table 3. A Comparison of Statistical Power in the Six Simulated Data Sets. The Theoretical Values of -2LL are Based on Analysis of the Theoretical Population Covariance Matrices. The Number of Twin-pairs is 9,000, and the Number of Replications is 1,000

DATA SET	MODEL 1		MODEL 2		MODEL 3
	ACE-AE	ACE-CE	ACE-AE	ACE-CE	ADE-AE
	MEAN -2 LL (SD)	MEAN -2 LL (SD)	MEAN -2 LL (SD)	MEAN -2 LL (SD)	MEAN -2 LL (SD)
Theoretical population covariance matrices	83.347 (18.204)	365.304 (38.200)	625.616 (50.005)	83.620 (18.234)	28.053 (10.498)
1. Non-censored continuous data	83.583 (17.712)	367.988 (35.632)	627.045 (45.754)	84.245 (17.342)	29.506 (10.787)
2. Non-censored discrete data (15 categories)	78.391 (17.017)	335.681 (33.895)	582.241 (44.798)	75.763 (16.447)	27.927 (10.486)
3. Censored discrete data (8 categories)	25.436 (12.881)	259.637 (42.549)	336.466 (47.447)	59.437 (20.792)	45.865 (16.588)
4. Square root transformed censored data	27.583 (12.751)	250.526 (39.616)	335.946 (43.172)	57.857 (19.375)	41.761 (15.283)
5. Normal scores of censored data	29.410 (12.735)	228.906 (35.763)	324.412 (41.969)	51.337 (17.443)	35.741 (13.786)
6. Categorical data (4 categories based on censored data)	23.298 (9.545)	84.15322 (18.951)	167.982 (27.309)	17.661 (11.028)	8.705 (5.707)

Note:-2 LL= minus 2 log likelihood; The true model parameters in Model 1: Va=.50, Vc=.20, Ve=.30; Model 2: Va^2=.20, Vc=.50, Ve=.30; Model 3: Va=.45, Vd=.25, Ve=.30

[a] In the categorical data analysis, the number of replications was 999, 999, and 996 for model 1 to 3 respectively.

Table 3 shows the results of the power analyses. We have used a type I error rate of .05. Because of the large sample size we are not interested in power *per se*, but in the effect of the estimation method on power. In Table 3, we first report the theoretical value of the difference in -2 LL and its standard deviation. The theoretical value of the difference in -2 LL was determined by analysing the population covariance matrices in Mx. It is equal to the number of degrees of freedom (df) plus the non-centrality parameter (λ). The standard deviation was calculated with the following formula: $SD=(2(df+2*\lambda))^{0.5}$. As can be seen in Table 3, the mean -2LL of the continuous data analyses was quite similar to the theoretical value of -2LL.

It is important to realize that the values of the non-centrality parameter can only be interpreted in terms of null and non-null distributions of the likelihood ratio test in the case of the normally distributed data (continuous or 15 point scale), and in the case of the categorical data estimator. For example, the results observed in the case of model 3 seem to suggest that the power increases after censoring (e.g., from 28.053 to 45.865). However, this is due to the fact that the test statistics do not follow their expected non-central and central chi-square distributions. This is also true in case of the transformed censored data.

After discretization, the mean difference in -2 LL decreased slightly. This is a reflection of the decreased power due to a loss of information after discretization. In all three models, the power was lowest when the categorical data were analysed. This is evident in the low mean difference in -2 LL. In addition, when we look at the categorical data analyses of the ADE model, the drop of the D parameter did not lead to a significantly worse fit in 20% of the cases, although this parameter explained 25% of the variance. In other words, the power to detect a dominant genetic parameter that explains 25% of the variance is 80%. In comparison, the power is 100% when any of the other methods of analysis is chosen.

Genetic analyses of sleep problems

To illustrate the previous findings, we analysed data on sleep problems in 6,375 MZ twins and 12,192 DZ twins. The descriptive statistics are summarized in Table 4. These descriptives are only reported for the first-borns to save space; the descriptive statistics of the second borns are similar. The skewness and kurtosis of the raw scores are similar to the skewness and kurtosis after censoring in the simulation study. A square root transformation and the computation of normal scores both reduced the skewness and kurtosis.

The correlations were computed in four different ways, namely the Pearson product moment correlation (ppmc) of the untransformed raw scores, the square root transformed scores, and the normal scores. In addition, polychoric correlations of the categorical data were computed. The estimates are shown in Table 5. The ppmc's were quite similar, but the correlations in the categorical data were somewhat different and were higher in both MZ and DZ twins.

Table 4. Descriptive Statistics of Maternal Child Behavior Checklist Reports on Sleep Problems in 9,415 Three-year-old Dutch Twins (first-borns only)

Data set	Mean	Standard Deviation	Skewness	Kurtosis
Raw data (15 categories)	1.983	2.229	1.444	5.220
Square root transformed data	1.083	.900	.165	1.990
Normal scores	2.074	2.053	.540	2.613

Table 5. Twin Correlations for Maternal Child Behavior Checklist Reports on Sleep Problems

Data set	Correlation MZ (N=3,162 complete pairs)	Correlation DZ (N=6,053 complete pairs)
Raw scores (ppmc)	.745	.384
Square root transformed data (ppmc)	.741	.408
Normal scores (ppmc)	.748	.406
Categorical data (pc)	.786	.451

Note:ppmc=Pearson product moment correlation; pc=polychoric correlation

Based on the correlations, an ACE model seemed to be most plausible. The MZ correlation was slightly less than twice the DZ correlation, which implies the absence of Vd and a small contribution of Vc. Genetic model fitting analyses of untransformed, transformed, and categorical data showed that the influences of A and C were both significant. Table 6 shows the point estimates and the confidence intervals of the standardized variance components (Va, Vc, and Ve) on sleep problems. The estimate of Va was similar across methods, which was to be expected in the light of the results of the simulation study. The estimate of Vc ranged from .055 to .081 when normal theory ML was used. The estimate was .116 in the categorical data analysis. In contrast, the estimate of Ve was lower in the categorical data analyses. As expected, the categorical data analyses showed wider confidence intervals than the analyses based on normal theory ML.

Table 6. Estimates and 95% Confidence Intervals of Standardized Estimates of Genes and Environment on Maternal Child Behavior Checklist Reports of Sleep Problems

Data set	Va (low-high)	Vc (low-high)	Ve (low-high)
Raw scores	.676 (.628-.724)	.055 (.011-.097)	.270 (.256-.284)
Square root transformed data	.658 (.611-.705)	.081 (.038-.122)	.262 (.248-.276)

Note:Va=proportion of variance explained by additive genetic effects; Vc= proportion of variance explained by common environmental effects; Ve= proportion of variance explained by unique environmental effects

Table 6. Estimates and 95% Confidence Intervals of Standardized Estimates of Genes and Environment on Maternal Child Behavior Checklist Reports of Sleep Problems

DATA SET	VA (LOW-HIGH)	VC (LOW-HIGH)	VE (LOW-HIGH)
Normal scores	.665 (.618-.712)	.077 (.034-.118)	.258 (.245-.272)
Categorical data	.675 (.620-.733)	.116 (.064-.166)	.210 (.194-.227)

Note:Va=proportion of variance explained by additive genetic effects; Vc= proportion of variance explained by common environmental effects; Ve= proportion of variance explained by unique environmental effects

DISCUSSION

This paper deals with the effects of censoring on parameter estimates and statistical power in genetic analyses of quantitative traits. The censoring of normal distributed data results in data with a L-shaped distribution. The distribution resembles the distribution of most behavioral checklist data. This paper looks at the effects of censoring through a series of simulations. Data were simulated in accordance with three genetic models: two ACE models with different factor loadings of A, C, and E, and one ADE model.

Multivariate normal data were simulated and discretized because behavior checklist data are usually discrete. We replicated the finding of Dolan (1994) that discretization of normal distributed data does not lead to biased parameter estimates when the number of categories is seven or more and when the distribution is symmetric. Next, the simulated data were censored, which resulted in L-shaped distributions. When analysing the censored ACE data with normal theory ML, the common environmental component was underestimated while the unique environmental component was overestimated. Transformation of the data did not eliminate this bias, although the skewness and kurtosis decreased. Interestingly, a common finding in behavioral genetic studies is a small influence of shared environment on individual differences in behavior. This may partly be due to the fact that the influence of this component is underestimated when L-shaped data are analysed with normal theory ML. However, the underestimation of the relative influence of the additive genetic component was only 8-10%. When analysing the ADE data, a quite large underestimation of about 20% of the additive genetic component was found and both D and E were overestimated by about 10%.

In order to examine if these results hold when a smaller percentage of the data is censored, we examined the amount of bias in parameter estimates when 10% of the data set was censored instead of 39%. In this situation, the bias decreased and ranged from 3% in the ACE model to 5% in the ADE model (data not shown). Thus, depending on the level of censoring, the results of twin studies which have used normal theory ML to analyse L-shaped distributed data may be biased.

The bias in parameter estimates may be avoided by using categorical data analysis. However, this method has three disadvantages. First, the statistical power is reduced. This result is in agreement with the results of the simulation study of Neale et al. (1994) who found that in categorical data analysis approximately three times the sample size was needed for equivalent power to continuous data analysis. In our study, the decrease in power was most apparent when the simulated ADE data were analysed. Even with 9,000 twin-pairs, the power to detect a dominant genetic component that explains 25% of the variation, decreased with 20% (from 100% to 80%). However, one should realize that the type-II error rate may be lower when censored data are analysed with normal theory ML compared to categorical data analysis, but that the actual type-I error rate may be higher than the hypothesized value of .05. One way to deal with the low power, is to choose a type I error rate of .10 or .15 instead of .05. Second, the analyses are computationally more demanding. This problem may be solved by using Weighted Least Squares (WLS) in Lisrel (Jöreskog & Sörbom, 1996a). However, this method has the disadvantage that missing data are excluded which can be a problem when dealing with incomplete twin-data or with lon-

gitudinal data in which observations may be missing at some time-points. A third disadvantage is that the contingency tables may have empty cells. One remedy to the presence of empty cells is to decrease the number of categories.

To illustrate the results of the simulation study, we analysed real-life data on sleep problems. The skewness and kurtosis were similar to the skewness and kurtosis of the simulated data. In this example, the small common environmental influence explained enough variance to be detected in the categorical data analysis. The heritability was quite stable and ranged from 66% to 68%. The estimate of the common environmental influence was somewhat higher in the categorical data analysis (12%) than in the other analyses (6 to 8%). Based on the results of the simulation study, we can conclude that the estimate of 12% in the categorical analysis, is the correct estimate. The unique environmental influence ranged from 21 to 27%. In conclusion, sleep problems are, like other behavioral problems in young children, explained by large genetic influences and moderate environmental influences. The latter include shared environmental influences.

The main question that we addressed was: what is the best approach when analysing L-shaped distributed phenotypic data? The results of the simulation study show that the analysis of L-shaped distributed data with normal theory ML is not advisable when the data show high skewness and kurtosis. Transformations may reduce the skewness and kurtosis but do not eliminate the bias in parameter estimates. Categorical data analysis is a better option, because this is the only method with which unbiased parameter estimates are obtained. Because this estimation method has its own limitations, the best option would be to develop checklists that measure variation in the whole latent distribution of behavior. To this end, items should reflect both well-adapted and dysfunctional behavior.

REFERENCE LIST

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Amos, C. I. (1994). Robust Variance-Components Approach for Assessing Genetic-Linkage in Pedigrees. *American Journal of Human Genetics, 54,* 535-543.

Bielak, L. F., Sheedy, P. F., & Peyser, P. A. (2001). Coronary artery calcification measured at electron-beam CT: Agreement in dual scan runs and change over time. *Radiology, 218,* 224-229.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Crittenden, L. B. (1961). An interpretation of familial aggregation based on multiple genetic and environmental factors. *Annals of the New York Academy of Sciences, 91,* 769-780.

Derks, E. M., Hudziak, J. J., Beijsterveldt van, C. E. M., Dolan, C. V., & Boomsma, D. I. (2004). A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. *Behavior Genetics, 34,* 571-583.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology, 47,* 309-326.

Epstein, M. P., Lin, X. H., & Boehnke, M. (2003). A tobit variance-component method for linkage analysis of censored trait data. *American Journal of Human Genetics, 72,* 611-620.

Falconer, D. S. (1965). Inheritance of Liability to Certain Diseases Estimated from Incidence Among Relatives. *Annals of Human Genetics, 29,* 51-71.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. (4 ed.) Harlow, UK: Pearson Education.

Hambleton, R. K. & Swaminathan, H. (1985). *Item Response Theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff.

Jöreskog, K. G. & Sörbom, D. (1996a). *LISREL 8: Structural Equation Modeling with the SIMPLIS command language*. Chicago, Il: Scientific Software International.

Jöreskog, K. G. & Sörbom, D. (1996b). *PRELIS 2 User's Reference Guide*. Chicago: Scientific Software International.

Jöreskog, K. G., Sörbom, D., Du Toit, S., & Du Toit, M. (1999). *LISREL 8: New statistical feautures*. Chicago, Il: Scientific Software International.
Koot, H. M., Oord van den, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1997). Behavioral and emotional problems in young preschoolers: cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *Journal of Abnormal Child Psychology, 25,* 183-196.
Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.
Muthén, B. O. & Muthén, L. (2001). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
Nagin, D. & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development, 70,* 1181-1196.
Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA: Department of Psychiatry.
Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publisher.
Neale, M. C., Eaves, L. J., & Kendler, K. S. (1994). The Power of the Classical Twin Study to Resolve Variation in Threshold Traits. *Behavior Genetics, 24,* 239-258.
Oord van den, E. J. C. G., Pickles, A., & Waldman, I. (2003). Normal variation and abnormality: an empirical study of the liability distributions underlying depression and delinquency. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44,* 180-192.
Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral Genetics*. New York.
Prescott, C. A. (2004). Using the Mplus computer program to estimate models for continuous and categorical data from twins. *Behavior Genetics, 34,* 17-40.
Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2003). Heritability of attention problems in children: I. Cross-sectional results from a study of twins, age 3-12 years. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 117B,* 102-113.
Rietveld, M. J. H., Valk van der, J. C., Bongers, I. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D. I. (2000). Classification of twin zygosity. *Twin Research, 3,* 134-141.
Tobin, J. (1958). Estimation of Relationships for Limited Dependent-Variables. *Econometrica, 26,* 24-36.
Venables, W. N., Smith, D. M., & R Development Core Team (2002). *An introduction to R*. Bristol, United Kingdom: Network Theory Limited.

CHAPTER

8

A TEST OF THE EQUAL ENVIRONMENT ASSUMPTION (EEA) IN MULTIVARIATE TWIN STUDIES

This paper is published as Eske M. Derks, Conor V. Dolan, Dorret I. Boomsma (2006). A test of the Equal Environment Assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics, 9,* 403-411.

ABSTRACT

In the classic twin design, estimation of genetic and environmental effects is based on the assumption that environmental influences are shared to the same extent by monozygitic (MZ) and dizygotic (DZ) twins (Equal Environment Assumption; EEA). We explore the conditions in which the EEA can be tested based on multivariate phenotypic data. We focus on the test whether the correlation between shared environmental factors in DZ twins (r_C) is less than 1. First, model identification was investigated analytically in Maple and Mx. Second, statistical power was examined in Mx. Third, the amount of bias caused by violation of the EEA was evaluated. Finally, applications to empirical data concern spatial ability in adolescents and aggression in children. Bivariate and trivariate models include several instances in which the EEA can be tested. The number of twin-pairs that is needed to detect violation of the EEA with a statistical power of .80 (= .05) varied between 508 and 3576 pairs for the situations considered. The bias in parameter estimates, given misspecification, ranged from 5-34% for additive genetic effects, and from 4-34% for shared environmental effects. Estimates of the non-shared environmental effects were not biased. The EEA was not violated for spatial ability or aggression. Multivariate data provide sufficient information to test the validity of the EEA. The number of twin pairs that is needed is no greater than the number typically available in most twin registries. The analysis of spatial ability and aggression indicated no detectable violation of the EEA.

INTRODUCTION

Genetic and environmental contributions to individual differences in phenotypic traits can be estimated using genetically informative data. In the case of twin data, the estimates are based on the monozygotic (MZ) and dizygotic (DZ) twin variances and covariances. With such data, a number of competing models can be considered. These models may include additive genetic effects (A), dominant genetic effects (D), shared environmental effects (C), and non-shared environmental effects (E). Estimates of these effects in the twin design are based, *inter alia,* on the following three assumptions: i) the additive (dominant) genetic effects correlate 1 (1) in MZ twin-pairs and .5 (.25) in DZ twin pairs, ii) the shared environmental effects correlate 1 in both MZ and DZ twin pairs, and iii) the non-shared environmental effects do not correlate in MZ and DZ twin pairs. In this paper, we explore the possibilities of testing the second assumption, usually referred to as the Equal Environment Assumption (EEA), given bivariate or trivariate data. The EEA implies that the shared environmental influences are equally important in MZ and DZ twin-pairs. The validity of the EEA has been debated (Joseph, 2000; Faraone & Biederman, 2000).

It has been shown that MZ twins in childhood more often share playmates, share the same room, and dress more alike than same-sex DZ twins (Loehlin & Nichols, 1976). However, this does not necessarily imply that the EEA is violated. Firstly, the greater environmental similarity in MZ than DZ twins does not have to be related to a greater phenotypic similarity. Secondly, even if a greater environmental similarity *is* related to a greater phenotypic similarity, this association could be mediated by a greater genetic similarity in MZ than DZ twins (Scarr & Carter-Saltzman, 1979). The EEA is only violated when the correlation between environmental similarity and trait similarity is significantly greater than zero *within* zygosity groups. Eaves et al. (2003) concluded on the basis of simulation studies that the absence of any association between environmental similarity and trait similarity justifies the claim that environmental similarity is not a factor in twin resemblance. However, the counter-claim that the presence of an association between environmental similarity and trait similarity falsifies the EEA is unfounded.

Several methods have been proposed to detect violation of the EEA. One way of assessing the validity of the EEA is to see whether *within* zygosity groups, differences in environmental experiences are associated with differences in phenotypic traits (Loehlin & Nichols, 1976). The influence of several environmental experiences have been studied, including ratings of treatment similarity (Loehlin & Nichols, 1976; Rowe et al., 2002; Wade et al., 2003), physical similarity (Hettema et al., 1995), and frequency of contact as adults (Kendler et al., 2000a). Loehlin & Nichols (1976) reported low correlations between differential treatment scores and mental ability, personality traits, vocational interests, and interpersonal relationships within zygosity groups. The range of correlations was -.15 to +.22, which is about what one would expect on the basis of change fluctuation. This suggests that the EEA is not violated for a wide variety of phenotypic traits. The validity of the EEA was also confirmed by more recent studies, which included measures of the influence of specified environmental measures on psychiatric diseases, such as major depression, generalized anxiety disorder, panic disorder, phobias, post-traumatic stress disorder, nicotine dependence, marijuana dependence, alcohol dependence, and psychoactive substance use and abuse (Kendler et al., 2000a; Hettema et al., 1995; Kendler & Gardner, 1998; Xian et al., 2000). The EEA has also been found to be tenable for sexual orientation (Kendler et al., 2000b). In contrast, possible violation of the EEA was reported for smoking initiation (Kendler & Gardner, 1998), and bulimia (Kendler & Gardner, 1998; Hettema et al., 1995; Rowe et al., 2002).

Another test for the validity of the EEA is provided by twins whose genetic similarity is misperceived by themselves and others (Scarr & Carter-Saltzman, 1979). If beliefs about zygosity determine the extent to which the two members of a twin-pair are behavior-

ally similar, then DZ twins who believe they are MZs should be similar to MZs. Likewise, MZs who believe they are DZ will be as different as true DZs. Scarr & Carter-Saltzman (1979) showed that twin resemblance with respect to cognitive measures was in accordance with the true, not the self-perceived, zygosity. For personality measures, DZs who believed they were MZs were more similar than those who correctly believed they were DZ. However, it turned out that beliefs about zygosity were highly related to genetic similarity at 12 loci, so the segregating genes of these DZ pairs might indeed correlate higher than .5. Recent studies suggest that perceived zygosity does not influence similarity on psychiatric and substance dependence disorders (Kendler & Gardner, 1998; Xian et al., 2000).

Finally, correlations of MZ twins reared together can be compared to correlations of MZ twins reared apart (MZA; Bouchard et al., 1990). MZA twins presumably do not share environmental influences so that the MZA correlation can be interpreted as an heritability estimate. Thus, a violation of the EEA has no consequence for the heritability estimate obtained with the MZA study design. Bouchard et al. (1990) report a heritability of about 70% for mental ability, and a heritability of about 50% for personality variables. As these heritability estimates resemble those based on data from MZ and DZ twins reared together, the EEA seems tenable.

The inclusion of specified familial environmental measures or perceived zygosity provides a means to detect violation of the EEA with respect to these environmental measures. However, a violation of the EEA will not be detected if it is related to factors other than those that are measured. For instance, the EEA was violated for some environmental measures related to bulimia (Hettema et al., 1995; Wade et al., 2003), but not for others (Kendler et al., 1991). Another concern with these methods is that they can not be applied in the absence of an environmental measure or index of perceived zygosity. The purpose of this paper is to explore an alternative approach to testing the validity of the EEA in multivariate data. Given multivariate data, and provided certain conditions are met, the shared environmental correlation can be estimated in DZ twins. If, in these situations, the shared environmental correlation in DZ twins does not deviate significantly from 1, this would suggest that the EEA is tenable. Below, we first identify the conditions in which estimation of the shared environmental correlation in DZ twins is possible. Next, we examine the statistical power to detect a violation of the EEA, and estimate the amount of bias (e.g., overestimation of the genetic effects and underestimation of the shared environmental effects) introduced by the given violation of the EEA. Finally, we apply this method to data on two indicators for spatial ability in 12-19 year old twins and three indicators of maternal rated aggression in 7 year old twins.

MATERIALS AND METHODS

Model description

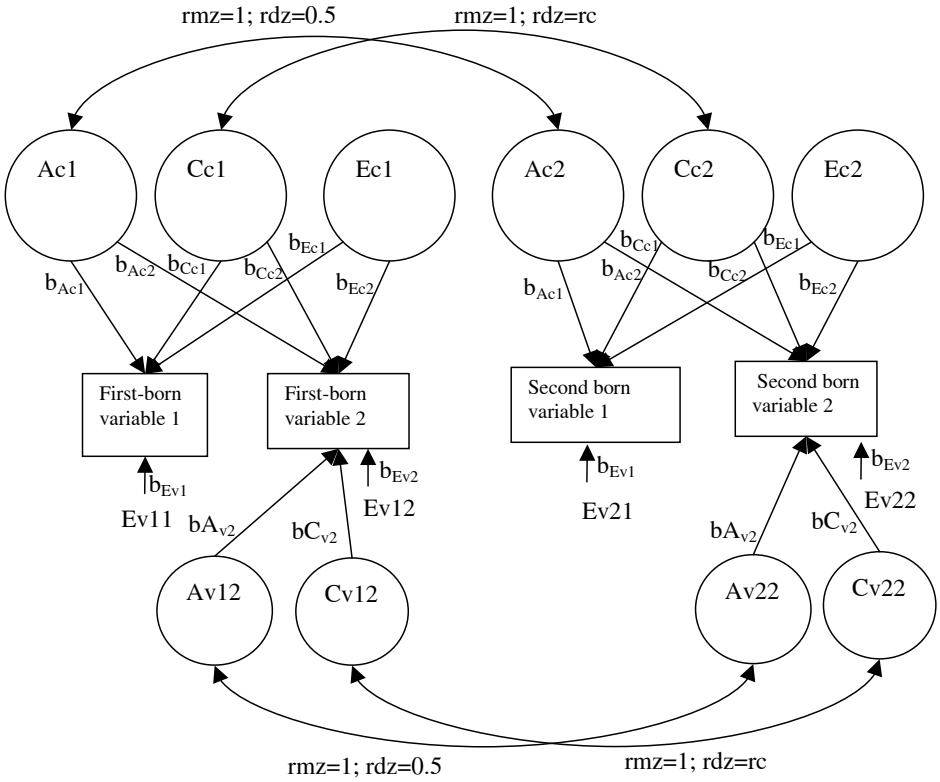
We employ a standard biometric model to test the EEA in bivariate and trivariate data sets. We assume that the phenotypic MZ and DZ covariance matrices are sufficient statistics as this facilitates the assessment of identifiability. Given m phenotypes (m=2 or 3 in the present paper), the number of observed statistics, i.e., (co)variances, is (nv*(nv+1), where nv equals m*2 (we do not consider the means, as they provide no information). However, not all variances and covariances have unique expectations in the twin model. Given m=2 and nv=4, the total number of observed statistics is 20. However, the number of statistics with unique expectations is nine, consisting of two variances (for variable 1 and variable 2), one within-subjects covariance, and six between subject's covariances (three in MZ twins and three in DZ twins). Given m=3 and nv=6, the total number of observed statistics in MZ and

DZ twins is 42. The number of statistics with unique expectations is 18, consisting of three variances, three within subjects covariances, and 12 between subjects covariances (six in MZ twins and six in DZ twins).

In the standard biometric model, additive genetic influences (A) are assumed to correlate 1 in MZ twins and .5 in DZ twins. The nonshared environmental influences (E) are uncorrelated in MZ and DZ twins. Influences of A, C and E may be common to all observed phenotypes (Ac, Cc, and Ec) or may be variable specific (Av, Cv, and Ev).

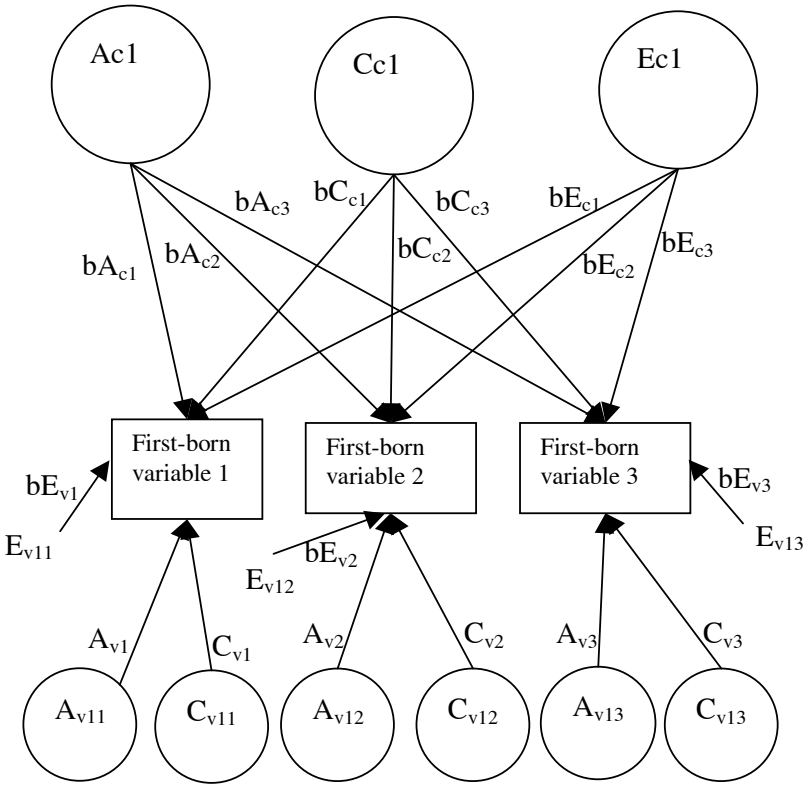
In the case of m=2, nine parameters are estimated in this biometric model, which results in a model that is just identified and no degree of freedom is left to test the validity of the EEA. Below, we explore the constraints that do allow such a test. After establishing that the various constraints do not lead to a decrease in model fit, the validity of the EEA can be tested by constraining the correlation of the Cc variables (r_{Cc}) at 1 in MZ twins, and estimating it freely in DZ twins. The correlation of the Cv variables, r_{Cv}, can be equated to r_{Cc} or fixed to 1. In this paper, we chose to equate r_{Cv} and r_{Cc} ($r_{Cv} = r_{Cc} = r_C$). If r_C is significantly lower than 1 in DZ twins, this suggests that the EEA is violated. Graphical representations of the biometric model for two and three observed variables are shown in Figure 1 and Figure 2, respectively.

Figure 1. Illustration of the biometric model to test the validity of the Equal Environment Assumption (EEA) with two observed variables



Note: Variation in variable 1 and 2 is explained by common (c), and variable specific (v) additive genetic effects (Ac, Av), shared environmental effects (Cc, Cv), and non-shared environmental effects (Ec, Ev). The correlation of A (ra) is 1 in MZ twins and .5 in DZ twins. The correlation of C (r_C) is 1 in MZ twins, and is freely estimated in DZ twins. In addition, the influence of Av can only be estimated if the influences of Ac are constrained to be equal for variable 1 and variable 2. The r_C of Cv should be equated to the r_C of Cc or should be constrained at 1. Although the model is not identified in its current form, constraints which render the model identified can be made. These constraints are provided in Table 2.

Figure 2. Illustration of the biometric model to test the validity of the Equal Environment Assumption (EEA) with three observed variables



Note: Variation in variable 1, 2, and 3 is explained by additive genetic effects (Ac, Av), shared environmental effects (Cc, Cv), and non-shared environmental effects (Ec, Ev). For the sake of clarity, the model is only given for one member of a twin-pair. The correlations of Ac and Av are 1 in MZ twins and .5 in DZ twins, and the correlations of Cc and Cv (r_C) are 1 in MZ twins, and freely estimated in DZ twins (but constrained to be equal for Cc and Cv). Although the model is not identified in its current form, eight different constraints which render the model identified can be made. These constraints are provided in Table 3.

Simulation analyses

We established model identification by calculating the null space of the Jacobian of the model. In contrast to the empirical test (necessary, but not sufficient), this test is necessary and sufficient (Bekker et al., 1993). Let s denote the q=nv*(nv+1) dimensional vector of expected (co)variances of both the MZs and the DZs, and let q denote the p-dimensional vector of free (i.e., unknown, to be estimated) parameters. The p x q Jacobian matrix simply equals $^1s/^1q$. As explained by Bekker et al. (1993), the model is identified if the null space is empty. We used Maple (e.g., Heck, 1993) to carry out this test (Maple worksheet is available on email request).

To establish if there are certain conditions in which r_C can not be estimated in DZ twins, expected (i.e., population) covariance matrices were calculated in Mx (Neale et al., 2003), given a choice of parameter values. The parameter values of the factor loadings ranged between -5 and 5, and the parameter values chosen for r_C ranged between .3 and .8. Biometric models were fit to these covariance matrices. DZ twins usually outnumber MZ twins, we therefore maintained a 1:2 ratio of MZ to DZ. Model identification was checked empirically by establishing that the true parameter values were recovered, and by comput-

ing the confidence intervals of the parameter estimates. If the model is identified, the true parameter values should be recovered exactly regardless of the starting values, the chi-square should equal zero, and the calculation of confidence intervals should pose no computational problems.

Having established model identification, we investigated the statistical power by calculating the number of twin-pairs that is required to reject the constrained model with r_C=1 with a probability of .80 and a significance level of .05. To estimate the bias caused by the violation of the EEA, we compared the simulated standardized genetic and environmental estimates with the estimated standardized genetic and environmental influences when the shared environmental correlation was constrained at 1 in DZ twins (i.e., misspecified).

We performed power calculations for one model with two observed variables, and for one model with three observed variables. In the model with two observed variables, the loadings of Ac on variable 1 and 2 equalled 1 and 2, respectively. The loadings of Cc on variable 1 and 2 equalled 2 and 1, respectively. The loadings of Ec, and Ev equalled 1 on both variables. The loadings of Av and Cv on variable 2 equalled 0. In the model with three observed variables, loadings of Ac equalled 1, 2, and 3; Av equalled 2, 2, and 2; Cc equalled 3, 2, and 1; and Cv equalled 2, 2, and 2, for variable 1, 2, and 3, respectively. Ec and Ev equalled 1 for all three observed variables.

Analysis of two indicators for spatial ability

Spatial ability data were available for 171 same-sex MZ twin pairs, and 133 same-sex DZ twin-pairs aged 12 to 19 years old (Osborne, 1980). We did not consider sex differences in view of the relatively small sample size. The two indicators for spatial ability are cube comparison and surface-development. In the cube comparison test, each item consists of two cubes of which 3 sides are visible. The testee has to determine whether the cubes are possibly identical. In the Surface-development test, the testee has to determine whether a piece of paper with a given form can be folded in to a given three dimensional form (e.g., a square box).

Analysis of three indicators for aggression

Mothers of 1534 Dutch twin-pairs who are registered with the Netherlands Twin Register (Boomsma et al., 2002) completed the Child Behavior Checklist (CBCL; Achenbach, 1991) and the Conners Parent Rating Scale-Revised: Short version (CPRS-R:S; Conners, 2001) when the children were seven years old. Subjects were divided into 6 groups: male MZ (244 pairs), male DZ (269 pairs), female MZ (285 pairs), female DZ (249 pairs), male–female opposite-sex pairs (first-born is male, second born is female; 243 pairs), and female-male opposite-sex pairs (first-born is female, second born is male; 244 pairs). The CBCL contains 20 items on aggression, which can be subdivided into two subscales: direct aggression (6 items) and relational aggression (14 items; Ligthart et al., 2005). The CPRS-R:S contains 6 items on oppositional behavior. Analyses were performed on sum scores of these three subscales: direct aggression, relational aggression, and oppositional behavior.

RESULTS

Model identification

The test based on the Jacobian and the empirical calculations pointed out that a number of bivariate and trivariate models allow estimation of r_C. The constraints that identify the model are summarized in Table 1 and Table 2, for two and three observed variables, respectively. It is possible that additional identifying constraits exist that were not considered in this paper. Before testing the validity of the EEA, it should first be established that the iden-

tifying constraint does not lead to a significant decrease in model fit. Otherwise, the possible misfit of the model in which the EEA is assumed to be tenable, can be the result of the misfit of the constraint.

Table 1. Overview of the constraints that allow estimation of the shared environmental correlation (r_C) in DZ twins for the bivariate model

	R_C (IN DZ TWINS)	IDENTIFYING CONSTRAINT	N PARAMETERS	IS THE MODEL IDENTIFIED?
Model 1	r_C =1	bEc1=bEc2 or bEv1=bEv2	9	Yes
Model 2	r_C =free	bEc1=bEc2 or bEv1=bEv2	10	No
Model 3	r_C =free	(bEc1=bEc2 or bEv1=bEv2) and bAc1=bAc2	9	Yes
Model 4	r_C =free	bEc1=bEc2 and bAv2=0	9	Yes

Note:
Model description (for explanation of the symbols used, please see Figure 1):
Twin 1, variable 1: $y11=b_{Ac1}*Ac1+b_{Cc1}*Cc1+b_{Ec1}*Ec1+b_{Ev1}*Ev11$
Twin 1, variable 2: $y12=b_{Ac2}*Ac1+b_{Cc2}*Cc1+b_{Ec2}*Ec1+b_{Av2}*Av12+b_{Cv2}*Cv12+b_{Ev2}*Ev12$
Twin 2, variable 1: $y21=b_{Ac1}*Ac2+b_{Cc1}*Cc2+b_{Ec1}*Ec2+b_{Ev1}*Ev21$
Twin 2, variable 2: $y22=b_{Ac2}*Ac2+b_{Cc2}*Cc2+b_{Ec2}*Ec2+b_{Av2}*Av22+b_{Cv2}*Cv22+b_{Ev2}*Ev22$

Table 2. Overview of the constraints that allow estimation of the shared environmental correlation (r_C) in DZ twins for the trivariate model

	R_C (IN DZ TWINS)	IDENTIFYING CONSTRAINT	N PARAMETERS	IS THE MODEL IDENTIFIED?
Model 1	r_C =1	None	18	Yes
Model 2	r_C =free	None	19	No
Model 3	r_C =free	bAv1=0 or bAv2=0 or bAv1=0	18	Yes
Model 4	r_C =free	bAv1=bAv2 or bAv1=bAv3 or bAv2=bAv3	18	Yes
Model 5	r_C =free	bAc1=bAc2 or bAc1=bAc3 or bAc2=bAc3	18	Yes
Model 6	r_C =free	bAc1=bCc1 or bAc2=bCc2 or bAc3=bCc3	18	Yes
Model 7	r_C =free	bEc1=bCc1 or bEc2=bCc2 or bEc3=bCc3	18	Yes
Model 8	r_C =free	bAv1=bCv1 or bAv2=bCv2 or bAv3=bCv3	18	Yes
Model 9	r_C =free	bAc1=bAv1 or bAc2=bAv2 or bAc3=bAv3	18	Yes
Model 10	r_C =free	bAc1=bEv1 or bAc2=bEv2 or bAc3=bEv3	18	Yes

Note:
Model description (for explanation of the symbols used, please see Figure 2):
Twin 1, variable 1: $y11=b_{Ac1}*Ac1+b_{Cc1}*Cc1+b_{Ec1}*Ec1+b_{Av1}*Av11+b_{Cv1}*Cv11+b_{Ev1}*Ev11$
Twin 1, variable 2: $y12=b_{Ac2}*Ac1+b_{Cc2}*Cc1+b_{Ec2}*Ec1+b_{Av2}*Av12+b_{Cv2}*Cv12+b_{Ev2}*Ev12$
Twin 1, variable 3: $y12=b_{Ac3}*Ac1+b_{Cc3}*Cc1+b_{Ec3}*Ec1+b_{Av3}*Av13+b_{Cv3}*Cv13+b_{Ev3}*Ev13$
Twin 2, variable 1: $y21=b_{Ac1}*Ac2+b_{Cc1}*Cc2+b_{Ec1}*Ec2+b_{Av1}*Av21+b_{Cv1}*Cv21+b_{Ev1}*Ev21$
Twin 2, variable 2: $y22=b_{Ac2}*Ac2+b_{Cc2}*Cc2+b_{Ec2}*Ec2+b_{Av2}*Av22+b_{Cv2}*Cv22+b_{Ev2}*Ev22$
Twin 2, variable 3: $y22=b_{Ac3}*Ac2+b_{Cc3}*Cc2+b_{Ec3}*Ec2+b_{Av3}*Av23+b_{Cv3}*Cv23+b_{Ev3}*Ev23$

The empirical calculations proved useful to explore specific configurations of parameter values, which rendered the model unidentified. First, the correlation of the shared environment (r_C) in DZ twins should not equal .50. Clearly, if r_C is equal to .50, the shared environmental effects cannot be distinguished from the genetic effects, as these also correlate 1 in MZ twins and .50 in DZ twins. Second, the factor loadings of Cc should not be identical for all observed variables. This second condition implies that the observed variables are not allowed to correlate perfectly. Third, the model is not identified if the factor loadings of Ac and Cc are collinear (i.e., the factor loadings of Ac and Cc on an observed

variable are the factor loadings of Ac and Cc on another variable, multiplied by a constant). We will illustrate the latter problem for the model with two observed variables V1 and V2 which are observed in twin 1 (T1) and twin 2 (T2):

I Covariance (V1T1,V1T2)=bA_{c1}*r_A*bA_{c1} + bC_{c1}* r_C *bC_{c1}

II Covariance (V1T1,V2T2)= bA_{c1}* r_A *bA_{c2}+ bC_{c1}* r_C*bC_{c2}

If, for a certain value for a constant c, $bA_{c2} = c^* bA_{c1}$, and $bC_{c2} = c^* bC_{c1}$, then

III) Covariance (V1T1,V2T2)=c*Covariance (V1T1,V1T2)

Therefore, it may seem that we observe four covariances (two covariances in MZ twins and two covariances in DZ twins), but two of the covariances are a function of the other two covariances plus the constant c, which results in only three pieces of unique information.

Statistical power to detect violation of the Equal Environment Assumption (EEA)

The statistical power to detect violation of the EEA depends on the magnitude of the genetic and shared environmental influences. For illustrative purposes, we performed power calculations for one model with two observed variables, and for one model with three observed variables. The parameter values of the models are provided in the methods section. The model with two indicators was identified by constraining the influences of Av and Cv on variable 2 at zero. If r_C equals .70, 618 twin-pairs are needed to detect violation of the EEA (for a statistical power of .80 at an of .05). If r_C equals .90, 3576 twin-pairs are needed to detect violation of the EEA. The model with three indicators was identified by constraining Av to be equal for variable 1, 2, and 3. In the case of r_C equals .70, 508 twin-pairs are needed to detect violation of the EEA. If r_C equals .90, 2111 twin-pairs are needed to detect violation of the EEA.

Bias in parameter estimates when violation of the Equal Environment Assumption is ignored

How large is the bias in parameter estimates when violation of the EEA is not accommodated in the fitted model? The true standardized influences of A, C, and E are compared to the estimated values when r_C is fixed at 1 in DZ twins for the two models that are described above. The influences of variable specific and common factors in the biometric model are summed. Table 3 summarizes the true and biased standardized estimates for the two models. Depending on the magnitude of the genetic and shared environmental influences, the heritability is overestimated with 5 to 34%, and the shared environmental influences are underestimated with 4 to 34%. The non-shared environmental influences do not show much bias.

Table 3. Bias in standardized parameter estimates if violation of the Equal Environment Assumption is not accommodated (true/biased estimate)

		Variable 1			Variable 2			Variable 3		
	r_C (DZ)	A	C	E	A	C	E	A	C	E
Model 1 (2 indicators)	.90	14/20	57/50	29/31	57/62	14/10	29/28	-	-	-
	.70	14/38	57/29	29/32	57/70	14/2	29/29	-	-	-

Note:A=additive genetic effects; C=shared environmental effects; E=non-shared environmental effects; r_C is the simulated value of the shared environmental correlation. The estimates of A, C, and E are summed over the common and variable specific factors. The true parameter estimates are obtained with r_C fixed at its true value. The biased parameter estimates are obtained with r_C constrained at 1.

Table 3. Bias in standardized parameter estimates if violation of the Equal Environment Assumption is not accommodated (true/biased estimate) *(Continued)*

		Variable 1			Variable 2			Variable 3		
	r_C (DZ)	A	C	E	A	C	E	A	C	E
Model 2 (3 indicators)	.90	25/35	65/54	10/10	44/55	44/34	11/11	65/70	25/20	10/10
	.70	25/58	65/31	10/10	44/78	44/11	11/11	65/81	25/9	10/10

Note:A=additive genetic effects; C=shared environmental effects; E=non-shared environmental effects; r_C is the simulated value of the shared environmental correlation. The estimates of A, C, and E are summed over the common and variable specific factors. The true parameter estimates are obtained with r_C fixed at its true value. The biased parameter estimates are obtained with r_C constrained at 1.

Is the EEA violated for spatial ability?

We tested the validity of the EEA by analyzing data on two indicators of spatial ability: cube comparison and surface-development. The results of the model fitting analyses are summarized in Table 4. As was shown before, the number of unique observed statistics in the variance/covariance matrix of these data is nine. First, an ACE model was fit to the data, in which a cholesky decomposition was specified for A, and C, and a common factor model for E (the full biometric model). This model estimates nine parameters for the covariance structure: 3 factorloadings for A, C, and E, respectively. Second, before testing the validity of the EEA, we had to find a constraint that identifies the model, but does not lead to a significant decrease in model fit. The loadings of Ac on cube comparison and surface-development were significantly different. Therefore, we could not identify the model by constraining these parameters to be equal. The loading of Av on surface-development was not significantly different from zero, so we identified the model in which the validity of the EEA can be investigated by constraining this loading at zero. Freely estimating r_C did not lead to significant increase in model fit, which suggests that the EEA is not violated. In Table 5, we included the parameter estimates of the best-fitting model.

Table 4. Model-fitting results and parameter estimates for two indicators of spatial ability in 304 twin pairs

Model	Constraints	-2 log LL	N Par	Compared to model	df	2	p
1. Full biometric model	bEc1=bEc2, r_C=1	8636.61	17*	-	-	-	-
2. Identifying constraint 1	bEc1=bEc2, bAc1=bAc2, r_C=1	8640.54	16	1	1	3.93	.047
3. Identifying constraint 2	**bEc1=bEc2, bAv2=0, r_C=1**	**8637.51**	**16**	**1**	**1**	**0.898**	**.343**
4. Violation of the EEA allowed	bEc1=bEc2, bAv2=0, rc=free	8636.61	17	3	1	0.898	.343

Note:The total number of parameters of the full biometric model=17, consisting of 8 parameters for the means and 9 parameters for the variance/covariance matrix. The best-fitting model is printed in bold.

Table 5. Overview of the standardized genetic and environmental influences on the variance and covariance of two indicators of spatial ability

Variable	A	C	E
Cube comparison	17	28	55
Surface development	59	16	24
Covariance cube comparison / surface development	58	39	3

Note:A=additive genetic effects, C=shared environmental effects, E=non-shared environmental effects

Is the EEA violated for aggression?

We tested the validity of the EEA with respect to aggression by analyzing three indicators for aggression: relational and direct aggression, and oppositional behavior. The model fitting results are shown in Table 6. First, we fitted a standard biometric model. The influ-

ences of Av on relational and direct aggression could be equated in both boys and girls, so we used this as the identifying constraint. Next, we fitted the model in which the r_C is estimated in DZ male twins and DZ female twins and opposite-sex twins. We tested if the r_C was significantly lower than 1 in DZ same-sex twins. We did not constrain the rC in opposite-sex twins at 1, because rC may be lower than 1 due to different environmental influences in boys and girls. The validity of the EEA was not violated. The parameter estimates of the best-fitting model are shown in Table 7.

Table 6. Model-fitting results and parameter estimates for three indicators of aggression in 1534 twin pairs

MODEL	CONSTRAINTS	-2 LOG LL	N PAR	COMPARED TO MODEL	DF	2	P
1. Full biometric model	rc=1	40845.61	73	-	-	-	-
2. Identifying constraint 1	**bAv1=bAv2 (in boys and girls)**	**40846.99**	**71**	**1**	**2**	**1.38**	**.502**
3. Violation of the EEA allowed	bAv1=bAv2, rc=free	40845.90	73	2	2	1.09	.579

Note:The total number of parameters of the full biometric model=73, consisting of 36 parameters for the means; and 37 parameters for the variance/covariance matrix (18 in boys, 18 in girls, and the shared environmental correlation in opposite-sex twins). The best-fitting model is printed in bold.

Table 7. Overview of standardized genetic and environmental influences on the variance and covariance of three indicators of aggression

VARIABLE	A GIRLS/BOYS	C GIRLS/BOYS	E GIRLS/BOYS
Relational aggression	63/44	7/37	30/18
Direct aggression	56/37	15/48	30/15
Oppositional	56/43	12/27	32/30
Covariance relational / direct aggression	73/34	5/55	22/11
Covariance relational aggression / oppositional	74/41	3/40	23/19
Covariance direct aggression / oppositional	68/35	2/46	30/19

Note:A=additive genetic effects, C=shared environmental effects, E=non-shared environmental effects

DISCUSSION

One of the most widely debated assumptions of the twin method is the Equal Environment Assumption (EEA). The EEA requires that *within* zygosity groups, differences in environmental experiences are not associated with differences in phenotypic traits. In this paper, we demonstrated that when two or more variables are observed in both members of a twin-pair, the shared environmental correlation can be estimated in DZ twins. By testing whether the shared environmental correlation is significantly lower than 1 in DZ twins, violation of the EEA can be detected in the absence of measured environmental variables.

We showed that estimation of the shared environmental correlation (r_C) is possible if: i) two or more indicators of a phenotypic trait are measured; ii) the shared environmental correlation in DZ twins is different from 0.5; iii) the factor loadings of Cc are not identical for all observed variables (which implies that the variables are not allowed to correlate perfectly); iv) the factor loadings of Ac and Cc are not collinear; and v) an identifying constraint that does not lead to a significant decrease in model fit exists. Condition ii does not have to be fulfilled when data from genetically unrelated siblings are available. In genetically unrelated siblings, additive genetic effects do not correlate and the correlation of the shared environmental effects is similar to the DZ shared environmental correlation.

If these five conditions are met, the validity of the EEA can be tested. Although it may seem that an additional requirement is that the influence of the shared environmental effects should be significantly greater than zero in the univariate analyses, this is not the case. In fact, the model is also identified when the shared environmental correlation is lower han .5 in DZ twins. In this case, the correlation of DZ twins is not more than half the MZ correlation and no significant influence of the shared environment would be found in the univariate analyses.

A concern when testing the validity of the EEA is that the correlation in opposite-sex twins may be lower than the MZ and same-sex DZ correlation as a result of different environmental influences in boys and girls. When combining data from opposite-sex and same-sex DZ twins, this may lead to rejection of the validity of the EEA while the lower correlation is actually the result of environmental sex-limitation. Therefore, in case of a significantly lower correlation in opposite-sex twins than in same-sex DZ twins, the data from same-sex DZ twins and opposite-sex twins should not be combined. The validity of the EEA can be tested by constraining the correlation of the shared environmental influences at 1 in same-sex DZ twins but not in opposite-sex twins.

The statistical power to detect violation of the EEA is acceptable. The number of twin pairs that is needed to detect violation of the EEA is no greater than the number of twin pairs typically available in most twin registries. Additional simulations showed that the power decreases when the factor loadings become more similar for the observed variables (data not shown). This is not surprising, as r_C can not be estimated when the observed variables correlate perfectly (i.e., the multivariate model simplifies to a univariate model). It was found that ignoring violation of the EEA can sometimes lead to large bias in parameter estimates. The influence of the additive genetic effects is overestimated (5-34%), and the influence of the shared environmental effects is underestimated (4-34%).

Analyses of empirical data showed that the EEA is not violated for spatial ability in adolescents or aggression in children. For both phenotypes, the shared environmental correlation was not significantly lower than 1 in DZ twins. This is in accordance with the results of previous studies in which the validity of the EEA was supported (Loehlin & Nichols, 1976; Scarr & Carter-Saltzman, 1979; Kendler et al., 2000a). It is reassuring that the EEA seems to be a valid assumption for most traits, although its validity should be examined whenever possible.

Although the proposed method is useful for testing the validity of the EEA when no environmental measures are available, it also has a number of limitations. First, if a trait is influenced by dominant genetic influences (e.g., Attention Deficit Hyperactivity Disorder), shared environmental influences cannot be included in the model and the shared environmental correlation can not be estimated. Therefore, in the presence of dominance, the proposed model cannot be used to test the validity of the EEA unless phenotypic data from other relatives are available. For example, inclusion of phenotypic data from the parents allows for the estimation of influences of A, C, D, and E on phenotypic variation if these parameters are not age dependent. A second concern is that the interpretation of possible violation of the EEA may be complicated. Imagine that a trait is influenced by a factor that correlates 1 in MZ twins and .7 in DZ twins. This factor can be interpreted as an environmental factor for which the EEA is violated. In contrast, it could also be a genetic factor, which correlates higher than .5 due to assortative mating. Therefore, given that $r_C<1$ in DZ twins, the EEA is not necessarily violated.

REFERENCE LIST

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Bekker, P. A., Merckens, A., & Wansbeek, T. J. (1993). *Indentification, equivalent models, and computer algebra.* Boston, MA: Academic Press.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of Human Psychological Differences - the Minnesota Study of Twins Reared Apart. *Science, 250,* 223-228.

Conners, C. K. (2001). *Conners' Rating Scales-Revised*. New York and Toronto: Multi-Health Systems, Inc.

Eaves, L., Foley, D., & Silberg, J. (2003). Has the "equal environments" assumption been tested in twin studies? *Twin Research, 6,* 486-489.

Faraone, S. V. & Biederman, J. (2000). Nature, nurture, and attention deficit hyperactivity disorder. *Developmental Review, 20,* 568-581.

Heck, A. (1993). *Introduction to Maple, A Computer Algebra System*. New York: Springer-Verlag.

Hettema, J. M., Neale, M. C., & Kendler, K. S. (1995). Physical Similarity and the Equal-Environment Assumption in Twin Studies of Psychiatric-Disorders. *Behavior Genetics, 25,* 327-335.

Joseph, J. (2000). Not in their genes: A critical view of the genetics of attention-deficit hyperactivity disorder. *Developmental Review, 20,* 539-567.

Kendler, K. S. & Gardner, C. O. (1998). Twin studies of adult psychiatric and substance dependence disorders: are they biased by differences in the environmental experiences of monozygotic and dizygotic twins in childhood and adolescence? *Psychological Medicine, 28,* 625-633.

Kendler, K. S., Karkowski, L. M., Neale, M. C., & Prescott, C. A. (2000a). Illicit psychoactive substance use, heavy use, abuse, and dependence in a US population-based sample of male twins. *Archives of General Psychiatry, 57,* 261-269.

Kendler, K. S., Maclean, C., Neale, M., Kessler, R., Heath, A., & Eaves, L. (1991). The Genetic Epidemiology of Bulimia-Nervosa. *American Journal of Psychiatry, 148,* 1627-1637.

Kendler, K. S., Thornton, L. M., Gilman, S. E., & Kessler, R. C. (2000b). Sexual orientation in a US national sample of twin and nontwin sibling pairs. *American Journal of Psychiatry, 157,* 1843-1846.

Ligthart, L., Bartels, M., Hoekstra, R. A., Hudziak, J. J., & Boomsma, D. I. (2005). Genetic contributions to subtypes of aggression. *Twin Research, 8,* 483-491.

Loehlin, J. & Nichols, R. (1976). *Heredity, Environment & Personality.* Austin & London: University of Texas Press.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA: Department of Psychiatry.

Osborne, R. T. (1980). *Twins: Black and White.* Athens: Foundation for Human Understanding.

Rowe, R., Pickles, A., Simonoff, E., Bulik, C. M., & Silberg, J. L. (2002). Bulimic symptoms in the Virginia twin study of adolescent behavioral development: Correlates, comorbidity, and genetics. *Biological Psychiatry, 51,* 172-182.

Scarr, S. & Carter-Saltzman, L. (1979). Twin Method - Defense of A Critical Assumption. *Behavior Genetics, 9,* 527-542.

Wade, T. D., Wilkinson, J., & Ben Tovim, D. (2003). The genetic epidemiology of body attitudes, the attitudinal component of body image in women. *Psychological Medicine, 33,* 1395-1405.

Xian, H., Scherrer, J. F., Eisen, S. A., True, W. R., Heath, A. C., Goldberg, J. et al. (2000). Self-reported zygosity and the equal-environments assumption for psychiatric disorders in the Vietnam Era Twin Registry. *Behavior Genetics, 30,* 303-310.

CHAPTER

9

A STUDY OF GENETIC AND ENVIRONMENTAL INFLUENCES ON MATERNAL AND PATERNAL CBCL SYNDROME SCORES IN A LARGE SAMPLE OF 3-YEAR-OLD DUTCH TWINS

This paper is published as: Derks E.M., Hudziak J.J., Beijsterveldt van C.E.M., Dolan C.V., Boomsma D.I. (2004). A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. *Behavior Genetics, 34,* 571-583

ABSTRACT

There is increasing evidence that behavioral problems are common in very young children, yet little is known about the etiology of individual differences in these problems. It is unclear to what degree environmental and genetic factors influence the development of early child psychopathology. In this paper we focus on the following issues: Firstly, to what degree do genetic and environmental factors influence variation in behavioral problems? Secondly, to what degree are these underlying etiological factors moderated by sex and informant? We investigate these issues by analyzing Child Behavior Checklist (CBCL) data on 9,689 3-year-old twin pairs. Rater Bias and Psychometric Models were fitted to CBCL/2-3 data obtained from mothers and fathers to determine the genetic and environmental contributions to the five CBCL syndromes: aggressive, oppositional, overactive, withdrawn, and anxious/depressed behavior. Parental ratings are influenced by aspects of the child's behavior that are experienced in the same way by both parents and by aspects of the child's behavior that are experienced uniquely by each parent. There is evidence for high genetic contributions to all CBCL syndromes. Shared and non-shared environmental influences play significant roles as well. One exception is overactive behavior, which is influenced by genetic and non-shared environmental influences only. Variation in behavior problems in the very young shows high heritability. Individual raters offer unique perspectives that can have an impact on estimates of problem behavior and genetic architecture. Therefore, multi-informant approaches in the assessment of the very young will be useful to clinicians and researchers alike.

INTRODUCTION

Little is known about the etiology of behavioral disorders in the very young. Studies in this age group have focused mainly on the assessment of problem behavior in clinically referred or at risk samples (Thomas & Guskin, 2001; Shaw et al., 2001). In order to obtain a better understanding of the etiology of psychopathology in non-clinical samples of very young children, we studied problem behavior in a sample of 9,821 3-year-old twin-pairs, using the Child Behavior Checklist (CBCL) (Achenbach, 1991; Achenbach, 1992).

The CBCL, completed by mothers and/or fathers, has been used in studies of childhood behavior around the world (Achenbach & Rescorla, 2001). Many studies have shown that CBCL-scores predict behavior problems in children as they age. For example, CBCL syndrome scores for overactive behavior at age 3 and attention problems at age 12 correlate .37 in a large Dutch twin sample (Rietveld et al., 2004). The correlation between aggressive scores at age 3 and 12 is .41 (Beijsterveldt van et al., 2003). Hofstra et al. (2000) identified children with deviant CBCL-scores, and showed that 41% of these children were still classified as deviant 14 years later. A final example that demonstrates the continuity of problem behavior is the study of Achenbach et al. (1995). CBCL-scores were used to predict symptoms of disturbance in adolescents, such as academic problems, suicidal behavior, and substance abuse. CBCL-scores accounted for an average of 31% of the variance of symptoms of disturbance measured three years later.

Parents and other caregivers are the main source of information when it comes to the assessment of problem behavior in young children. Because studies on the very young are relatively scarce, little is known about how best to use such parental reports. Fathers and mothers may not agree about the nature and degree of problem behavior in their children. Different raters are confronted with different aspects of their children's behavior, and may have different internal standards for the evaluation of behavior. For example, when DSM-interview data were collected from both parents to determine the presence or absence of psychopathology in their child, the correlations between maternal and paternal data ranged from .13 to .35 (Hewitt et al., 1997). Achenbach & Rescorla (2001) reported maternal and paternal correlations ratings of CBCL scales in 3-year-old twins that range from .48 to .67, with a mean correlation of .61. While these correlations are substantial, they still suggest that each parent has a unique perspective on the behavior of their offspring.

One advantage of using a twin population to study behavior problems, is that genetic models can be used to test whether maternal and paternal ratings diverge because of different internal standards (rater bias), or if their ratings reflect different, but valid, aspects of the child's behavior. In other words, when data from multiple informants are available, a distinction can be made between: a) variance explained by the environment that is shared between siblings, b) variance that is explained by rater bias, c) variance that is explained by a common perception of the parents, and d) variance that is explained by a unique perception of each parent. To investigate if the ratings of multiple informants disagree because of different internal standards or because of the reflection of different, but valid, perceptions of the child's behavior, two structural models have been developed. The Rater Bias Model (Hewitt et al., 1992; Neale & Stevenson, 1989) allows the parental ratings to be influenced by the behavior of the child and by rater bias, which gives rise to disagreement between the parents. The Psychometric Model (Hewitt et al., 1992) allows the parental ratings to be influenced by aspects of the child's behavior that are experienced commonly by both parents, and by aspects of the child's behavior that are experienced uniquely by each parent.

The genetic contributions to the two broadband scales (externalizing and internalizing) of the CBCL in Dutch 3-year-olds, as well as to the seven behavioral syndromes (aggressive, oppositional, overactive, withdrawn, anxious/depressed, sleep problems, and somatic problems) were reported by Valk van der et al. (1998; 2001) and by Oord van den et

al. (1996). Valk van der et al. (1998) studied maternal reports of externalizing and internalizing behavior in 3-year-old children, and found that additive genetic factors explained 54% of the variance in externalizing behavior, and 64% of the variance in internalizing behavior. Shared and non-shared environmental factors explained the remaining part of the variance of the CBCL broadband scales. More recently, Valk van der et al. (2001) combined maternal and paternal scores of externalizing and internalizing in 3-year-olds and compared the fit of the Rater Bias and Psychometric Model. The Psychometric Model fitted the data better than the Rater Bias Model. The parents were found to assess a common component in the behavior in their children, and, in addition, each parent experienced unique aspects of their children's behavior. The heritability of internalizing behavior was 66%. The heritability of 66% was mostly explained by a common view of the parents (86%), and in addition by a unique view of the parents (14%). The heritability of externalizing behavior was 54%. The common view of the parents explained 87% of the heritability and the unique view explained 13% (Valk van der et al., 2001). These results agree with the results of Hewitt et al. (1992), who studied internalizing behavior in an 8- to 11-year-old sample and in a 12- to 16-year-old sample of twins. The Psychometric Model fitted the data better than the Rater Bias Model in both samples. This supports the hypothesis that also at older ages, disagreement is not caused by parental bias, but by the fact that the parents assess partially different aspects of the children's behavior.

Additive genetic and environmental influences on the seven CBCL syndromes were reported by Oord van den et al. (1996) in a sample of 1,358 twin pairs. For all scales, parental ratings were for the most part expressions of the same underlying trait. Therefore, Oord van den et al. averaged parental scores in the genetic model fitting to obtain estimates of genetic and environmental influences. Additive genetic effects explained the main part of the variance of the CBCL syndromes (60% to 74%). Shared environmental effects influenced individual differences in oppositional and aggressive behavior (both 12%). Non-shared environmental influences explained the remaining part of the variance (19% to 40%). The authors acknowledged the possible influence of non-additive genetic effects on overactive behavior, but these effects were not included within the genetic model, due to inadequate power to detect these effects.

In this paper we analyze behavioral problems in over 9,600 Dutch twin-pairs aged 3 years. Ratings of their behavior were obtained from mothers and fathers. Because of the large sample size, we have sufficient statistical power to detect genetic dominance, and sibling interaction effects, if present. In addition, we will investigate whether the parameter estimates of the genetic model fitting are similar in boys and girls. Oord van den et al. found small sex differences, but did not include these in the final models. We analyze maternal and paternal ratings of aggressive, oppositional, overactive, withdrawn, and anxious/depressed behavior, because these scales represent the most common behavioral problems in the very young. To determine whether parents assess different aspects of their children's behavior, we fit Rater Bias and Psychometric Models. We focus on the behavioral syndromes instead of the broadband scales because the behavioral syndromes may form a better basis for prescribing treatment (Achenbach & Edelbrock, 1984), and may be more suitable for future gene-finding studies (Hudziak, 1997; Hudziak, 2001).

METHOD

SUBJECTS

This study is part of an ongoing longitudinal twin study in the Netherlands. The subjects were all registered at birth with the Netherlands Twin Registry (Boomsma et al., 2002). For the present study, we analyzed data of a sample of Dutch twins, whose parents (or primary caregivers) reported on their behavior when they were three years old. These twins were all

born between 1986 and 1997. Of the total sample of 9,969 pairs, 152 pairs were excluded from statistical analyses because a health questionnaire completed at age three indicated that one or both of the twins suffered from a disease or handicap that interfered severely with daily functioning. The resulting sample comprises 9,817 pairs. The zygosity status of 128 pairs was unknown; these pairs were excluded from the genetic analyses. The sample that was used for the genetic analyses consisted of 9,689 pairs.

Zygosity diagnosis was assessed with the use of a 10-item questionnaire. This procedure allows an accurate determination of zygosity of nearly 95%. It is described in more detail in (Rietveld et al., 2000). The number of twin pairs, by sex, zygosity and informant are presented in Table 1.

Table 1. Number of Twin Pairs by Sex, Zygosity and Informant

ZYGOSITY	FATHER	MOTHER	TOTAL
MZm	1033	1519	1561
DZm	1059	1594	1635
MZf	1156	1736	1777
DZf	972	1454	1494
DOS	2030	3142	3222
Unknown	76	125	128
Total zygosity known	6250	9445	9689
Total	6326	9570	9817

MZm=Monozygotic male, DZm=Dizygotic male
MZf=Monozygotic female, DZf=Dizygotic female
DOS=Opposite sex

PROCEDURE

A survey, including the CBCL/2-3, was mailed to the fathers and mothers of the twins when the twins were three years old. Due to funding problems, this questionnaire was only sent to the mother of the twins born between May 1989 and November 1991. Parents who did not return the forms within two months received a reminder, and during some years, persistent non-responders were contacted by phone four months after the initial mailing. This procedure resulted in a 75.5% participation rate (Rietveld et al., 2004).

MEASURE

The Child Behavior Checklist (CBCL/2-3) is a standardized questionnaire for parents to report the frequency and intensity of behavioral and emotional problems exhibited by their child in the past six months. It contains 100 items that measure problem behavior; the items are rated on a 3-point scale ranging from "not true", "somewhat or sometimes true", to "very true or often true". The CBCL measures the number of symptoms of seven behavioral syndromes, which can be combined to form two broadband scales: externalizing and internalizing behavior. The seven syndromes were derived from factor analyses of the problem items. These factor analyses resulted in the formation of the problem scales oppositional (17 items), withdrawn (10 items), aggressive (9 items), anxious/depressed (9 items), overactive (5 items), sleep problems (7 items) and somatic problems (3 items) (Koot et al., 1997). Sleep problems and somatic problems were not analyzed in this study, because the prevalence of these problems was very low.

STATISTICAL ANALYSES

Means and standard deviations of untransformed CBCL-scores were calculated using SPSS/Windows 11.0. (SPSS, 2001). The distributions of these scores are skewed. In order to obtain a distribution that approaches normality with respect to skewness and kurtosis, normal scores were computed with Prelis (Jöreskog & Sörbom, 1996).

The effects of sex and zygosity on these normal CBCL-scores were tested in an ANOVA. The effects were examined in first and second born twins separately, to avoid dependency inherent in twin data. The type-I error rate was corrected for multiple testing in two ways. First, the (type I error probability) of each test was set to equal .01. Second, an effect was only considered to be present if it was significant given is .01, in both first and second born twins.

Twin correlations among CBCL syndrome scores were computed in Mx (Neale et al., 2003). Structural Equation Modeling (SEM) was employed to obtain an estimate of the genetic and environmental contributions to the observed variances and covariances. An assumption of SEM is that the data are normally distributed. Therefore, these analyses were carried out on the normal scores (Jöreskog & Sörbom, 1996). The genetic model fitting analyses were performed on raw data with Mx, using maximum likelihood estimation. Point estimates and confidence intervals for the estimated genetic and environmental parameters are reported (Neale et al., 2003; Neale & Miller, 1997). Technical details of genetic model-fitting analyses are reviewed elsewhere (Neale & Cardon, 1992).

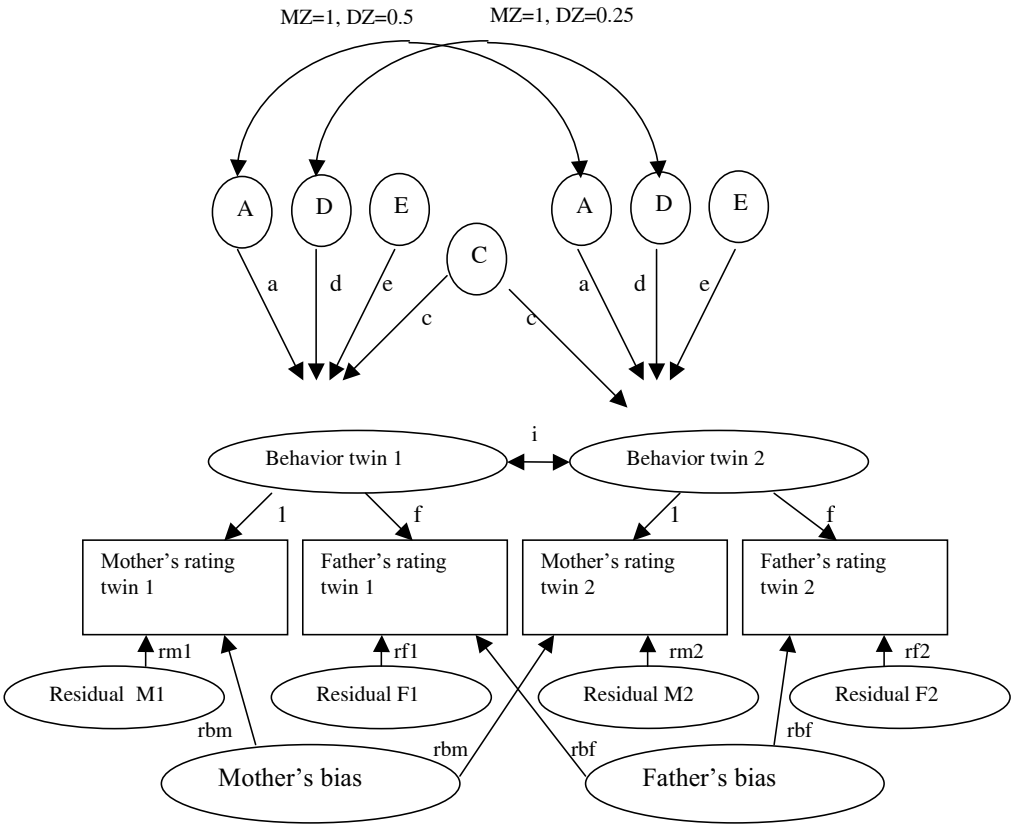
Model Fitting

Variation in a phenotypic trait can be decomposed into latent genetic and environmental components. The decomposition of variance takes place by comparing the degree of similarity between pairs of individuals, who differ in their degree of genetic relatedness. Monozygotic twins are genetically identical, while dizygotic twins on average share half of their segregating genes. Limiting the genetic decomposition of phenotypic variance to additive genetic (A) effects and dominant genetic (D) effects, the fact that MZ twins are genetically identical implies that they share all the additive genetic and dominant genetic variance. DZ twins on average share half of the additive genetic and one quarter of the dominant genetic variance (Plomin et al., 2001). In addition to the genetic components, the phenotypic variance is decomposed into shared and non-shared environmental variance. The shared environmental variance is due to environmental effects shared by two members of a twin pair (C). These effects are by definition perfectly correlated in both monozygotic and dizygotic twins. The non-shared environmental variance is due to effects (E) which are by definition uncorrelated between twin pair members. Estimates of the non-shared environmental variance usually include measurement error (Plomin et al., 2001). In fitting models to MZ and DZ twin data, it is not possible to estimate the effects of all mentioned sources of variance (A, D, C, and E). Specifically, with E and A in the model, one cannot estimate D and C simultaneously.

The Rater Bias Model (Figure 1) allows one to estimate variance due the effects of genetic and environmental factors (note that Figure 1 includes all 4 sources of variance A, D, C, E, even though, as mentioned, they cannot all be estimated simultaneously). In this model, the parental ratings of their children's behavior are not only influenced by the child's behavior, but also by rater bias and residual error. The influence of the child's behavior on the ratings of the fathers and the mothers may differ. To identify this model, the factor loading of the child's behavior on the maternal ratings is fixed to 1, whereas the factor loading of the child's behavior on the paternal ratings is freely estimated. In addition, in the full model, we do not constrain the parental bias to be equal for MZ and DZ twins. This allows for the possibility that parental biases are influenced by the beliefs that parents have about their twin's zygosity (Neale and Cardon, 1992).

The Psychometric Model (Figure 2; again including all 4 sources of variance) allows the parental ratings to be influenced by aspects of the child's behavior that are perceived commonly by both parents, and by aspects of the child's behavior that are perceived uniquely by each parent. Unique perceptions could arise if the child behaves differentially towards his or her parents, or if the parents observe the child in different situations. The common and unique aspects are both influenced by genetic and environmental factors.

Figure 1. Rater Bias Model



Note: A=additive genetic factor; D=dominant genetic factor; C=shared environmental factor; E=non-shared environmental factor; f=factorloading from behavior of twins on father rating

In both models we have added a path with coefficient i between the CBCL-scores of the twins. This path implies an interaction that may be interpreted in two ways (Simonoff et al., 1998). First, it may be considered an interaction between siblings (Carey, 1986; Eaves, 1976). Second, the path may be considered an effect introduced by the rater, who may compare the behavior of one child with the behavior of the other child. The latter may thus be interpreted as a rater contrast effect. Very low DZ correlations compared to MZ correlations give a first indication that a competitive social interaction effect or negative rater contrast effect is present. However, such a configuration of twin correlations also suggests an ADE model. One way to distinguish between these possibilities is by testing the observed variances for MZ and DZ twins. An interaction effect leads to different variances in MZ and DZ twins (Hewitt et al., 1992). In case of dominance, MZ and DZ variances are expected to be equal. If a cooperative social interaction effect or a positive rater contrast effect is present, the pattern of MZ and DZ correlations resembles an ACE model (i.e., DZ correlation greater than half the MZ correlation). Again the model that includes an interaction term gives rise to differences in variances of MZ and DZ twins (Eaves, 1976), and thus may be distinguished from an ACE model by comparing variances.

Model Fitting Procedure

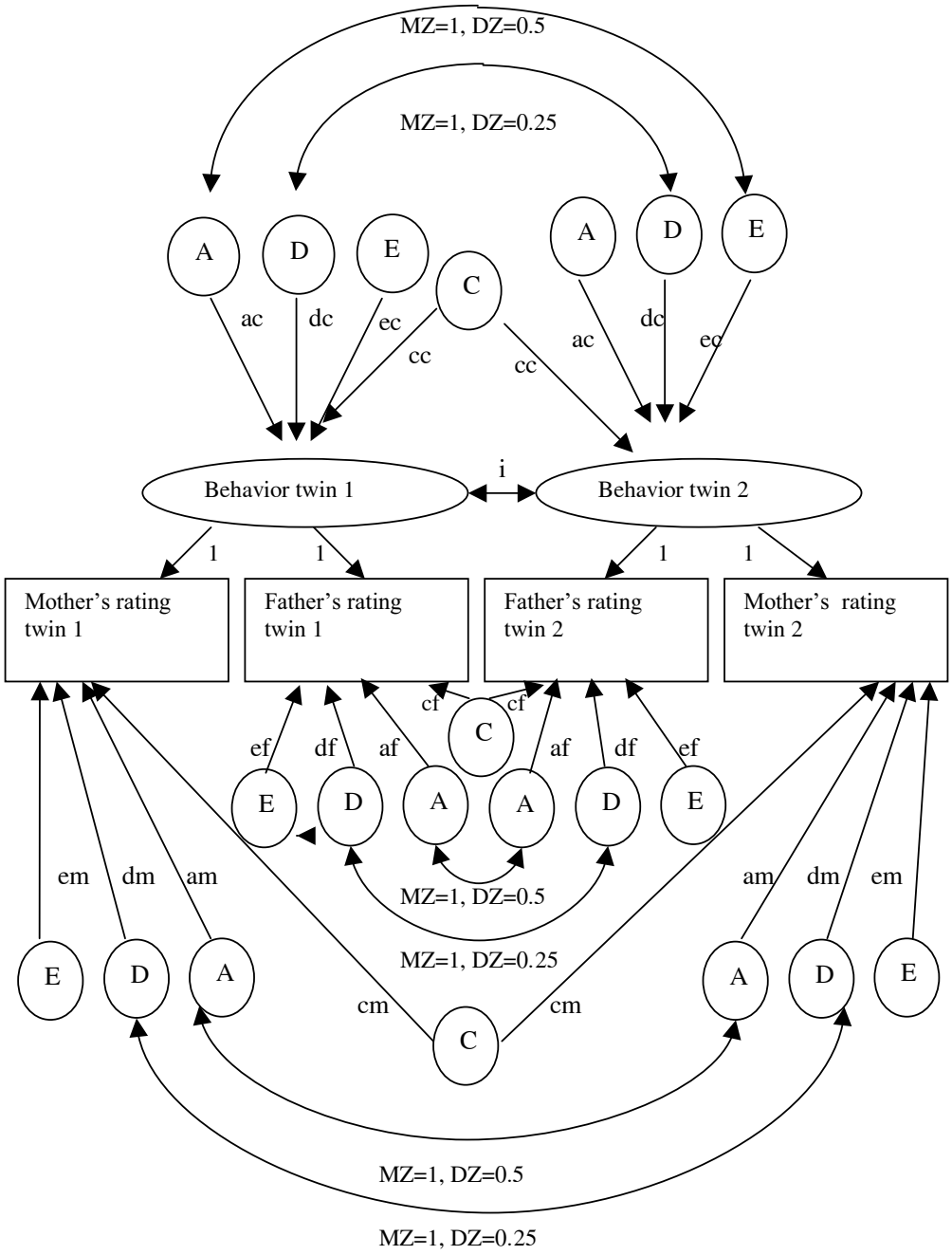
The first step in the model fitting procedure was to determine whether the interaction parameters were required (i.e., deviated significantly from zero). The fit of the model, in which the variances of MZ twins and DZ twins were constrained to be equal, was

compared with the fit of a fully saturated model, in which all variances and covariances were freely estimated. An interaction effect was included only if the variances of MZ and DZ twins were significantly different.

The second step was to choose between an ACE or an ADE model as the starting model. As mentioned above, with data of twins reared together, the effects of dominance and shared environment cannot be estimated simultaneously. The choice between the initial models was based on the phenotypic correlations: an ADE model was chosen if the MZ correlations were more than twice the DZ correlations, an ACE model was chosen if the MZ correlations were twice or less than twice the DZ correlations. The fit of the Rater Bias and Psychometric Model was then assessed by comparing the likelihood of these models to the likelihood of a fully saturated model.

In order to test for sex differences, we compared the likelihood of a model that includes estimates of parameters that vary over boys and girls with the likelihood of a model that equates all model parameters over sex. This test is also sensitive for absolute variance differences between boys and girls, because the absolute factor loadings were equated. Finally, the significance of the common and unique influences of A, and C or D was tested by means of likelihood ratio tests.

Figure 2. Psychometric Model



Note: A=additive genetic factor; D=dominant genetic factor; C=shared environmental factor; E=non-shared environmental factor; ac=additive genetic common; dc=dominant genetic common; ec=non-shared environment common; cc=shared environment common; am=additive genetic maternal; dm=dominant genetic maternal; em=non-shared environment maternal; cm=shared environment maternal; af=additive genetic father; df=dominant genetic father; ef=non-shared environment father; cf=shared environment father

RESULTS

The means and standard deviations of the five CBCL-scales are reported by the six different zygosity groups by maternal and paternal report in Table 2. Note that the final two columns in Table 2 give the means by sex.

Table 2. Mean CBCL-scores and standard deviations

	Informant	MZm	DZm	MZf	DZf	DOSm	DOSf	All Boys	All Girls
Oppositional	M	10.7 (6.5)	10.5 (6.6)	10.3 (6.5)	10.3 (6.5)	9.9 (6.5)	9.7 (6.3)	10.3 (6.5)	10.1 (6.5)
	F	10.0 (6.4)	9.9 (6.2)	9.5 (6.2)	9.7 (6.3)	9.3 (6.2)	9.1 (6.0)	9.8 (6.3)	9.4 (6.2)
Aggressive	M	4.3 (3.0)	3.9 (3.0)	2.9 (2.5)	2.8 (2.5)	3.1 (2.7)	3.0 (2.6)	4.0 (3.0)	2.7 (2.5)
	F	3.7 (2.9)	3.6 (2.8)	2.5 (2.3)	2.4 (2.3)	2.8 (2.5)	2.7 (2.5)	3.6 (2.8)	2.4 (2.3)
Overactive	M	3.0 (2.2)	2.8 (2.3)	2.6 (2.1)	2.6 (2.2)	2.6 (2.2)	2.5 (2.2)	2.9 (2.2)	2.5 (2.1)
	F	2.9 (2.2)	2.8 (2.2)	2.6 (2.1)	2.6 (2.2)	2.5 (2.2)	2.4 (2.1)	2.8 (2.2)	2.5 (2.1)
Withdrawn	M	1.2 (1.6)	1.2 (1.6)	1.1 (1.4)	1.2 (1.5)	1.1 (1.5)	1.0 (1.4)	1.2 (1.6)	1.1 (1.5)
	F	1.1 (1.5)	1.2 (1.5)	1.0 (1.3)	1.1 (1.5)	1.0 (1.4)	1.0 (1.4)	1.1 (1.5)	1.0 (1.4)
Anxious	M	3.5 (3.0)	3.5 (3.1)	3.8 (3.2)	3.8 (3.2)	3.4 (3.0)	3.2 (3.0)	3.5 (3.1)	3.5 (3.1)
	F	3.2 (2.8)	3.4 (3.0)	3.5 (3.0)	3.6 (3.0)	3.2 (2.9)	3.1 (2.9)	3.3 (3.0)	3.3 (3.0)

Estimates of skewness and kurtosis before and after normalization in Prelis are shown in Table 3. As expected, the transformation resulted in better (i.e., more normal) values of skewness and kurtosis. The transformed data were used for the tests of sex and zygosity effects, and for the genetic model fitting.

Table 3. Skewness and kurtosis of CBCL scales and their standard errors (SE) before and after transformation in Prelis

		Before Transformation		After Transformation (normal scores)	
		Skewness (SE)	Kurtosis (SE)	Skewness (SE)	Kurtosis (SE)
Maternal	Aggression	1.11 (.02)	1.27 (.04)	.15 (.02)	-.33 (.04)
	Oppositional	.59 (.02)	-.08 (.04)	.03 (.02)	-.13 (.04)
	Overactive	.67 (.02)	-.17 (.04)	.20 (.02)	-.47 (.04)
	Withdrawn	2.39 (.02)	8.51 (.04)	.56 (.02)	-.50 (.04)
	Anxious/Depressed	1.04 (.02)	.93 (.04)	.18 (.02)	-.42 (.04)
Paternal	Aggression	1.16 (.02)	1.47 (.04)	.18 (.02)	-.38 (.04)
	Oppositional	.62 (.02)	.03 (.04)	.04 (.02)	-.14 (.04)
	Overactive	.68 (.02)	-.15 (.04)	.21 (.02)	-.47 (.04)
	Withdrawn	2.23 (.02)	7.12 (.04)	.60 (.02)	-.49 (.04)
	Anxious/Depressed	1.03 (.02)	.92 (.04)	.20 (.02)	-.44 (.04)

Contributions of sex and zygosity on CBCL syndrome scores were tested in a two-way ANOVA. According to reports from both parents, boys had higher scores than girls on aggression, overactive, and withdrawn. Both maternal and paternal reports revealed higher scores in MZ than DZ twins on aggression and overactive behavior. Table 4 provides an overview of statistically significant main effects of sex and zygosity. No significant interaction effects of sex and zygosity were found. The inter-parent correlations of the CBCL syndromes are high and significant (Table 5). The correlations ranged from .54 to .71 for the five syndromes.

Table 4. An overview of the effects of sex and zygosity on transformed CBCL-scores

	Mother		Father	
	Sex	Zygosity	Sex	Zygosity
Aggressive	**	**	**	**
Oppositional				
Overactive	**	**	**	*
Withdrawn	*			
Anxious				

* p<.01** p<.001

Table 5. Parental correlations on transformed CBCL-scores

	Parental Correlation boys	Parental Correlation girls	Parental Correlation Opposite sex twins
Aggressive	.68	.66	.63
Oppositional	.71	.70	.71
Overactive	.68	.67	.66
Withdrawn	.53	.54	.55
Anxious	.68	.69	.68

Genetic analyses

The correlations of the twins' CBCL-scores are shown in Table 6. Based on the correlations, the ACE model served as the initial model for the genetic analyses on aggressive, oppositional, withdrawn and anxious/depressed behavior. The ADE model served as the initial model for overactive behavior.

The results of the genetic analyses are summarized in Table 7 for externalizing problem behaviors (aggressive, oppositional and overactive), and in Table 8 for internalizing problems (withdrawn and anxious/depressed). The best fitting model is printed in bold.

Tests of differences in variance between MZ and DZ twins revealed no significant differences. Therefore, interaction parameters were not included. For all syndromes, the Psychometric Model provided a better fit than the Rater Bias Model. The significance of A, and C or D was therefore tested by dropping A, and C or D from the Psychometric Model. Four significance tests were performed by fixing common influences of A, unique influences of A, common influences of C or D, and unique influences of C or D to zero. The significance of single parameters can be evaluated by considering the confidence intervals of the parameter estimates, which are reported for all parameters in the full Psychometric Model (Table 9).

Table 6. Maternal, paternal and cross-informant CBCL twin-correlations by zygosity status

		MZM	DZM	MZF	DZF	DOS-MF	DOS-FM
Aggression	Maternal	.83	.55	.83	.51	.48	.53
	Paternal	.80	.53	.83	.51	.51	.55
	Cross-rater	.57	.38	.60	.34	.30	.34
Oppositional	Maternal	.79	.53	.79	.50	.50	.51
	Paternal	.80	.57	.81	.54	.54	.57
	Cross-rater	.60	.38	.57	.35	.34	.38

Table 6. Maternal, paternal and cross-informant CBCL twin-correlations by zygosity status *(Continued)*

		MZM	DZM	MZF	DZF	DOS-MF	DOS-FM
Overactive	Maternal	.69	.14	.69	.15	.22	.19
	Paternal	.64	.12	.67	.21	.22	.18
	Cross-rater	.46	.04	.46	.08	.10	.08
Withdrawn	Maternal	.74	.46	.72	.51	.45	.42
	Paternal	.75	.50	.72	.55	.47	.49
	Cross-rater	.45	.27	.43	.34	.27	.28
Anxiety	Maternal	.68	.32	.72	.33	.38	.32
	Paternal	.67	.35	.71	.43	.39	.36
	Cross-rater	.49	.21	.53	.26	.24	.23

Table 7. Model fitting results, externalizing disorders

				Aggression (ACE)			Oppositional (ACE)			Overactive (ADE)		
	N PAR	With model	ΔDF	-2 LL	$\Delta\chi^2$	p	-2 LL	$\Delta\chi^2$	p	-2 LL	$\Delta\chi^2$	p
1. Fully saturated	84	-	-	132867.7	-	-	186356.3	-	-	125599.6	-	-
2. Equal variances [1]	76	1	8	132876.7	9.0	.34	186370.5	14.2	.08	125611.6	12.0	.15
3. Rater bias, unequal bias MZ/DZ	44	1	40	133109.7	242.0	.00	186469.3	113.0	.00	125825.8	226.2	.00
4. Rater bias, equal bias MZ/DZ	40	3	4	133125.5	15.8	.00	186487.0	17.7	.00	125871.3	45.6	.00
5. Psychometric + sex differences	42	1	42	**132923.5**	**55.7**	**.08**	186405.3	49.0	.21	125687.3	87.8	.00
6. Psychometric - sex differences	33	5	9	133024.3	100.9	.00	**186409.8**	**4.5**	**.48**	125702.7	15.3	.08
7. Psychometric unique A excluded[2]	38/31[2]	5/6[3]	4/2	133160.5	237.1	.00	186520.0	110.3	.00	125808.8	106.1	.00
8. Psychometric common A excluded[2]	40/32[2]	5/6[3]	2/1	133574.4	651.0	.00	187006.9	597.2	.00	125702.7	.0	-
9. Psychometric unique C/D excluded[2]	38/31[2]	5/6[3]	4/2	133054.7	131.2	.00	186748.1	338.3	.00	**125704.6**	**1.9**	**.38**
10. Psychometric common C/D excluded[2]	40/32[2]	5/6[3]	2/1	132977.7	54.27	.00	186470.5	60.8	.00	125918.5	215.8	.00

[1] The variances of MZ and DZ twins were fixed on similar values, to test if social interaction is plausible
[2] The number of parameters varies because sex differences were included if these were significant in previous model, they were excluded if non-significant in previous models.
[3] The model is compared to the psychometric model with sex differences if sex differences were significant and to the psychometric model without sex differences if sex differences were not significant

Table 8. Model fitting results, internalizing disorders

				Withdrawn (ACE)			Anxiety (ACE)		
	N par	With model	ΔDF	-2 LL	$\Delta\chi^2$	p	-2 LL	$\Delta\chi^2$	p
1. Fully saturated	84	-	-	102032.5	-	-	143968.7	-	-
2. Equal variances [1]	76	1	8	102041.5	9.0	.34	143983.1	14.5	.07
3. Rater bias, unequal bias MZ/DZ	44	1	40	102214.1	181.6	.00	144110.3	141.7	.00
4. Rater bias, equal bias MZ/DZ	40	3	4	102221.5	7.4	.12	144120.7	10.4	.03
5. Psychometric + sex differences	42	1	42	**102090.1**	**57.6**	**.06**	144025.9	57.2	.06
6. Psychometric - sex differences	33	5	9	102120.7	30.7	.00	144037.1	11.2	.26
7. Psychometric unique A excluded[2]	38/31[2]	5/6[3]	4/2	102242.3	152.2	.00	144183.6	146.5	.00
8. Psychometric common A excluded[2]	40/32[2]	5/6[3]	2/1	102344.3	254.2	.00	144558.6	521.5	.00
9. Psychometric unique C/D excluded[2]	38/31[2]	5/6[3]	4/2	102146.7	56.6	.00	144089.5	52.4	.00
10. Psychometric common C/D excluded[2]	40/32[2]	5/6[3]	2/1	102137.9	47.8	.00	**144037.1**	**.0**	-

[1] The variances of MZ and DZ twins were fixed on similar values, to test if social interaction is plausible
[2] The number of parameters varies because sex differences were included if these were significant in previous model, they were excluded if non-significant in previous models.
[3] The model is compared to the psychometric model with sex differences if sex differences were significant and to the psychometric model without sex differences if sex differences were not significant

Table 9. The standardized and unstandardized estimates of the genetic and environmental influences

		Common part			Unique part Paternal		Unique part Maternal	
		Stand. Paternal	Stand. Maternal	Non-Stand. (95% conf. interval)	Stand.	Non-Stand. (95% conf. interval)	Stand.	Non-Stand. (95% conf. interval)
Aggression Girls	A	50	44	2.98 (2.67-3.24)	11	.66 (.66-.88)	20	1.32 (1.31-1.62)
	C	10	9	.62 (.38-.70)	11	.67 (.60-.95)	9	.61 (.33-.61)
	E	8	7	.46 (.45-.53)	10	.59 (.58-.59)	11	.72 (.64-.80)
Aggression Boys	A	43	38	2.97 (2.97-3.39)	7	.51 (.51-.82)	19	1.52 (1.51-1.86)
	C	18	16	1.26 (1.26-1.69)	11	.77 (.49-.77)	10	.75 (.64-.76)
	E	11	10	.76 (.76-.86)	10	.68 (.64-.77)	8	.61 (.57-.70)
Oppositional Boys / Girls	A	19	18	18.16 (16.73-19.62)	4	.51 (.50-1.49)	15	5.60 (5.60-6.54)
	C	38	35	5.70 (5.68-7.11)	16	6.80 (5.93-7.75)	8	3.92 (3.92-5.01)
	E	16	15	4.90 (4.89-5.29)	17	2.97 (2.78-2.97)	9	4.06 (3.71-4.42)
Overactive Boys / Girls	A	0	0	.00 (.00-.08)	17	.80 (.80-.80)	18	.862 (.57-1.09)
	D	49	48	2.28 (2.16-2.39)	0	.00 (.00-.11)	4	.21 (.00-.48)
	E	19	19	.90 (.88-.97)	14	.67 (.67-.73)	11	.54 (.49-.54)
Withdrawn Girls	A	27	24	.53 (.39-.68)	13	.25 (.25-.38)	25	.56 (.56-.67)
	C	20	18	.40 (.26-.48)	14	.27 (.16-.40)	6	.13 (.13-.13)
	E	9	8	.18 (.15-.19)	18	.35 (.33-.39)	19	.41 (.41-.41)

Note:Stand. = standardized estimate of the variance
Nonstand. = non-standardized estimate of the variance

Table 9. The standardized and unstandardized estimates of the genetic and environmental influences *(Continued)*

		Common part			Unique part Paternal		Unique part Maternal	
		Stand. Paternal	Stand. Maternal	Non-Stand. (95% conf. interval)	Stand.	Non-Stand. (95% conf. interval)	Stand.	Non-Stand. (95% conf. interval)
Withdrawn Boys	A	40	36	.85 (.85-.95)	15	.32 (.18-.32)	24	.56 (.56-.56)
	C	8	8	.18 (.10-.28)	12	.26 (.16-.39)	7	.17 (.04-.32)
	E	10	9	.21 (.21-.24)	15	.33 (.33-.38)	18	.42 (.42-.47)
Anxiety Boys / Girls	A	54	49	4.64 (4.41-4.82)	5	.42 (.42-.62)	20	1.93 (1.86-2.10)
	C	0	0	.00 (.00-.13)	11	.95 (.94-1.22)	1	.06 (.00-.36)
	E	18	16	1.54 (1.42-1.65)	13	1.10 (.99-1.22)	14	1.29 (1.18-1.41)

Note:Stand. = standardized estimate of the variance
Nonstand. = non-standardized estimate of the variance

The variance in aggressive behavior is explained adequately by additive genetic influences (A), shared environmental influences (C), and non-shared environmental influences (E) ($\chi^2(42)=55.74$, p=.08). These factors were mainly explained by a common view of the parents (about 65%), but the factors that were viewed uniquely by each parent also explained a significant part of the variance. The influence of the common and unique factors differed significantly in boys and girls ($\chi^2(9)=100.86$, p=.00). The estimates show that the genetic influences are larger in girls than in boys while the shared environmental influences are higher in boys than in girls. The variance in oppositional behavior is explained by common and unique influences of A, C, and E ($\chi^2(51)=53.44$, p=.38). The influence of these factors are identical in boys and girls (χ^2 (9)=4.48, p=.48). In comparison with the other behavior scales, large shared environmental influences were found. The best fitting model for overactive behavior included A (common and unique), D (common only), and E (common and unique) ($\chi^2(53) = 105.03$, p = .00). Although common influences of A were non-significant, this factor was not removed from the model, because the presence of dominant influences in the absence of additive genetic influences is biologically implausible (Falconer & Mackay, 1996). The influences of A, D, and E were similar in boys and girls (χ^2 (9)=15.33, p=.08). The fit of the best-fitting model is rather poor (p< .05). However, it should be remembered that the χ^2 fit index is a function of sample size, which in the present case is large (N=9,689 pairs).

The variance in withdrawn behavior is explained by an ACE model (χ^2 (42)=57.61, p=.06). Both common and unique factors are present, but the common factors were slightly more important than the unique factors. The effects of these factors were significantly different in boys and girls (χ^2 (9)=30.65, p=.00). The genetic influences were higher in boys than in girls while the shared environmental influences were higher in girls than in boys. Anxious/depressed behavior in children is influenced by A (common and unique), C (unique only), and E (common and unique) (χ^2 (52)=68.46, p=.06). These effects were similar for boys and girls (χ^2 (9)=11.23, p=.26).

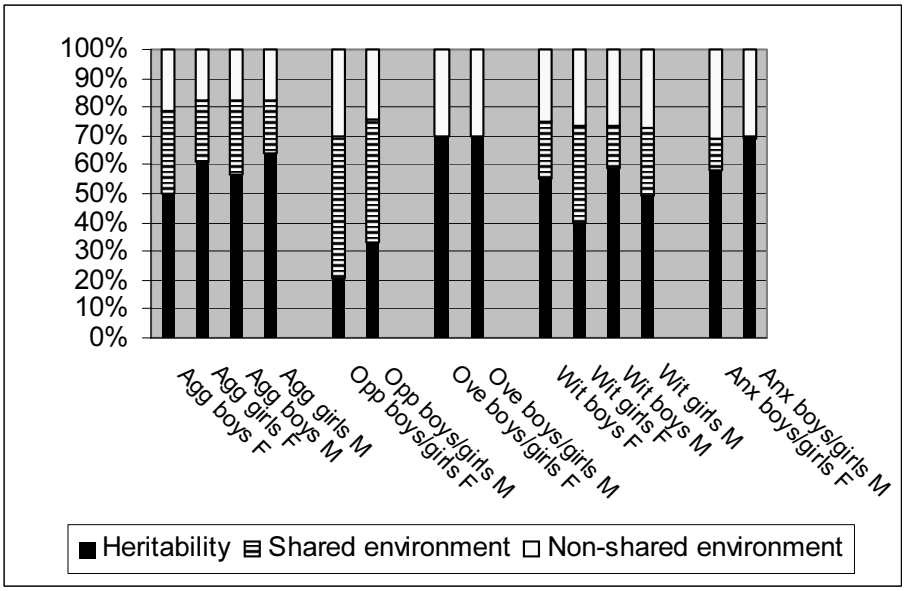
Table 9 reports the standardized and unstandardized estimates of the genetic and environmental influences. For example, aggressive behavior is explained by factors that influence both parental ratings (common view), and by factors that influence only the paternal or only the maternal rating (unique view). The paternal ratings of girls are explained by these common factors for 50 (A) + 10 (C) + 8 (E) = 68%. The remaining variance is unique to the paternal rating: 11 (A) + 11 (C) + 10 (E) = 32%. The heritability of the paternal ratings of aggression in girls can be calculated by summing the standardized variances that are explained by the genetic factor that influences both parental ratings (50%)

and the genetic factor that influences the paternal rating only (11%). The heritability of the paternal ratings is therefore 61%. Likewise, the heritability of the maternal ratings of aggression in girls is 64%.

The lack of significant sex differences in oppositional, overactive, and anxiety implies that the absolute variances are also equal in boys and girls. The sex differences in aggression and withdrawn may be caused by differences in the relative influences of A, C, and E but also by differences in absolute variances. Additional tests revealed that the absolute variances in aggression were slightly higher in boys than in girls. The absolute variances in withdrawn behavior were not significantly different between boys and girls.

A graphical representation of the genetic and environmental influences on behavioral problems is given in Figure 3. The genetic influences are the sum of common and unique additive and dominant genetic influences and is therefore a representation of the total heritability. The shared environmental influences are the sum of the common and unique shared environmental influences. Likewise, the non-shared environmental influences are the sum of the common and unique non-shared environmental influences.

Figure 3. Figure 3Graphical representation of the influences of genes, and shared and non-shared environment (common + unique influences)



Note: Agg=aggression, Opp=oppositional, Ove=overactive, Wit=withdrawn, Anx=anxious/depressed; F=father, M=mother

DISCUSSION

The goal of this study was to examine the relative influences of genes and environment on variation in problem behavior in 3-year-old boys and girls. We choose the CBCL because it is a widely used, quantitative, highly reliable instrument.

The data show that mothers and fathers agree to a large extent about the degree of problem behavior in 3-year-old children. The correlations we report are slightly higher than those reported for preschool children by Achenbach (1992) and Koot (1993), but similar to the correlations found in older children (Achenbach, 1991). For all scales, the Psychometric Model provided a better fit to the data than the Rater Bias Model. This implies that differences between parental reports are not only influenced by rater bias, but by aspects of the child's behavior that are perceived uniquely by each parent. This is in agreement with the

findings of Hewitt et al. (1992) and Valk van der et al. (2001). Although the parental ratings were influenced by unique perceptions of the child's behavior, the major part of the variance in problem behavior was explained by aspects of the child's behavior that were perceived commonly by the parents. These common perceptions explained 50 to 73% of the variance in the problem behavior scales.

Individual differences in problem behavior in 3-year-old children are mainly due to genetic differences. The large sample size allowed us to test whether shared environment contributes to problem behavior. To date it has been difficult to determine whether the often reported absence of shared environmental influences is due to the actual absence of these influences or to inadequate power to detect them in the classical twin design (Rutter et al., 1999). Confidence intervals are usually wide, even with a sample size as large as 2,682 twin pairs (Slutske et al., 1997). With the current sample size of 9,689 Dutch 3-year-old twin pairs, we detected significant shared environmental influences on four of the five scales: aggressive, oppositional, withdrawn and anxious/depressed. Because multiple rater data have been used, these shared environmental influences are not confounded with rater bias. However, as is to be expected in view of previous failures to detect these effects, the percentage of variance explained by shared environment was low compared to the percentage of variance explained by genes. The low DZ correlations in overactive behavior suggest that the presence of shared environmental influences on this problem scale is unlikely, although its influence could not be tested formally, due to inclusion of dominant genetic effects in the model. Thus, genetic effects are the most important etiological factor in problem behavior in young children, although shared and non-shared environmental influences are also present.

The present finding of large genetic influences on behavior in 3-year-olds suggests that the results of other studies, which do not take genetic effects into account, may be misinterpreted. For example, Carter et al. (2001), found that children of mothers, who experience a depressive disorder in addition to anxious behavior, substance abuse or an eating disorder, are at high risk for attachment insecurity. According to Carter et al. (2001) this attachment insecurity is caused by a less optimal interaction pattern of the depressed mothers. The results of the present study show that it is likely that the children of depressed mothers show similar symptoms because of the genes they received from their mothers or because of an interaction between these environmental and genetic factors.

An extensive literature exists on the presence of sex differences in psychopathology (for a review, see Rutter et al., 2003). However, sex comparisons are often based on specialized clinic-groups rather than on representative general population samples. In the present study, sex differences were examined in a large general population (twin) sample. Sex differences were found on aggressive, overactive, and withdrawn behavior. On these three scales, the scores of the boys were higher than those of the girls. The findings on aggressive and overactive behavior are consistent with the perception that boys show more of these behaviors than girls. Indeed, similar differences in scores on aggressive behavior have been reported at ages 7, 10, and 12 (Beijsterveldt van et al., 2003). The finding of higher maternal scores in boys than girls on withdrawn behavior is unexpected, but the size of the effect does not seem to be of clinical significance.

Sex differences in relative importance of genetic and environmental influences on individual differences were found in aggressive and withdrawn behavior, but not in oppositional, overactive, and anxious/depressed behavior. Compared to boys, individual differences in aggressive behavior in girls were influenced more by genes, and less by shared environment. In contrast, compared to boys, individual differences in withdrawn behavior in girls were more influenced by shared environment and less by genes.

The results of the present analyses of parental data on 9,689 3-year-old twin-pairs show that behavioral syndromes of early childhood are primarily influenced by genetic factors. Additive genetic factors account for the majority of these influences in all syndromes except for the parental ratings of overactive behavior, where dominant genetic factors were found to be more important. Non-shared and shared environmental effects also contribute to the expression of the common syndromes of early childhood problem behavior. The contribution of shared environment at this early age is plausible and expected. It will be interesting to determine whether shared environmental influences increase or decrease as the children age.

Parental reports were found to be influenced mainly by aspects of the child's behavior that are perceived commonly by the parents. However, parents also report on aspects of the child's behavior that are experienced uniquely by each parent. These unique aspects may arise because the child behaves differently towards both parents, or because both parents observe the child in different situations.

The finding of relatively large genetic contributions to early child psychopathology may facilitate gene finding expeditions. Specifically the finding that individual differences in behavioral problems are largely attributable to genetic influences increase the likelihood that chromosomal areas will be found to contribute to the phenotypic variance in linkage analyses. These results also have implications for diagnostics. The presence of heritable influences this early in life implies that the diagnosis of behavioral problems in young children should take into account a possible (early) history of behavioral problems in the parents. Needless to say, environmental factors cannot not be discarded, even in the presence of established familial history of behavioral problems.

REFERENCE LIST

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. & Edelbrock, C. S. (1984). Psychopathology of Childhood. *Annual Review of Psychology, 35,* 227-256.

Achenbach, T. M., Howell, C. T., McConaughy, S. H., & Stanger, C. (1995). Six-year predictors of problems in a national sample. III. Transitions to young adult syndromes. *Journal of the American Academy of Child and Adolescent Psychiatry, 34,* 658-669.

Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.

Beijsterveldt van, C. E. M., Bartels, M., Hudziak, J. J., & Boomsma, D. I. (2003). Causes of stability of aggression from early childhood to adolescence: A longitudinal genetic analysis in Dutch twins. *Behavior Genetics, 33,* 591-605.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior genetics, 16,* 319-340.

Carter, A. s., Garrity-Rokous, F. E., Chazan-Cohen, R., Little, C., & Briggs-Gowan, M. J. (2001). Maternal depression and comorbidity: Predicting early parenting, attachment security, and toddler social-emotional problems and competencies. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 18-26.

Eaves, L. J. (1976). A model for sibling effects in man. *Heredity, 36,* 205-214.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. (4 ed.) Harlow, UK: Pearson Education.

Hewitt, J. K., Silberg, J. L., Neale, M. C., Eaves L.J., & Erickson, M. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics, 22,* 293-317.

Hewitt, J. K., Silberg, J. L., Rutter, M., Simonoff, E., Meyer, J. M., Maes, H. et al. (1997). Genetics and developmental psychopathology .1. Phenotypic assessment in the Virginia twin study of adolescent behavioral development. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38,* 943-963.

Hofstra, M. B., Ende van der, J., & Verhulst, F. C. (2000). Continuity and change of psychopathology from childhood into adulthood: a 14-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 850-858.

Hudziak, J. J. (1997). Identification of phenotypes for molecular genetic studies of common childhood psychopathology. In K.Blum & E. Noble (Eds.), *Handbook of Psychiatric Genetics* (pp. 201-217). Boca Raton: CRC Press.

Hudziak, J. J. (2001). The role of phenotypes (diagnoses) in genetic studies of attention-deficit/hyperactivity disorder and related child psychopathology. *Child and Adolescent Psychiatric Clinics of North America, 10,* 279-97, viii.

Jöreskog, K. G. & Sörbom, D. (1996). *PRELIS 2 User's Reference Guide*. Chicago: Scientific Software International.

Koot, H. M. (1993). *Problem Behaviour in Dutch preschoolers*. Erasmus University, Rotterdam.

Koot, H. M., Oord van den, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1997). Behavioral and emotional problems in young preschoolers: cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *Journal of Abnormal Child Psychology, 25,* 183-196.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA: Department of Psychiatry.

Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publisher.

Neale, M. C. & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics, 27,* 113-120.

Neale, M. C. & Stevenson, J. (1989). Rater Bias in the EASI temperament scales: a twin study. *Journal of Personality and Social Psychology, 56,* 446-455.

Oord van den, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1996). A genetic study of maternal and paternal ratings of problem behaviors in 3-year-old twins. *Journal of Abnormal Psychology, 105,* 349-357.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral Genetics*. New York.

Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2004). Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 577-588.

Rietveld, M. J. H., Valk van der, J. C., Bongers, I. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D. I. (2000). Classification of twin zygosity. *Twin Research, 3,* 134-141.

Rutter, M., Caspi, A., & Moffitt, T. E. (2003). Using sex differences in psychopathology to study causal mechanisms: Unifying issues and research strategies. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44,* 1092-1115.

Rutter, M., Silberg, J., O'Connor, T., & Simonoff, E. (1999). Genetics and child psychiatry: I Advances in quantitative amd molecular genetics. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40,* 3-18.

Shaw, D. S., Owens, E. B., Giovanelli, J., & Winslow, E. B. (2001). Infant and toddler pathways leading to early externalizing disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 36-42.

Simonoff, E., Pickles, A., Hervas, A., Silberg, J. L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine, 28,* 825-837.

Slutske, W. S., Heath, A. C., Dinwiddie, S. H., Madden, P. A. F., Bucholz, K. K., Dunne, M. P. et al. (1997). Modeling genetic and environmental influences in the etiology of conduct disorder: A study of 2,682 adult twin pairs. *Journal of Abnormal Psychology, 106,* 266-279.

SPSS (2001). *SPS for Windows, Rel 11.0.1*. Chicago, Ill: SPSS.

Thomas, J. M. & Guskin, K. A. (2001). Disruptive behavior in children: what does it mean? *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 44-51.

Valk van der, J. C., Oord van den, E. J. C. G., 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 and Allied Disciplines, 42,* 921-931.

Valk van der, J. C., Verhulst, F. C., Stroet, T. M., & Boomsma, D. I. (1998). Quantitative genetic analysis of internalising and externalising problems in a large sample of 3-year-old twins. *Twin Research, 1,* 25-33.

CHAPTER

10

GENETIC ANALYSES OF MATERNAL AND TEACHER RATINGS ON ATTENTION PROBLEMS IN SEVEN-YEAR-OLD DUTCH TWINS

This paper is published as Derks E.M., Hudziak J.J., Beijsterveldt van C.E.M., Dolan C.V., Boomsma D.I. (in press). Genetic analyses of maternal and teacher ratings on Attention Problems in seven-year-old Dutch twins. *Behavior Genetics*.

ABSTRACT

The goal of the present study is to examine genetic and environmental influences on maternal and teacher ratings of Attention Problems (AP) in 7-year-old children. Teachers completed the Teacher Report Form (N=2259 pairs), and mothers the Child Behavior Checklist (N=2057 pairs). Higher correlations were found in twins rated by the same teacher than in twins rated by different teachers. This can be explained by rater bias or by a greater environmental sharing in twins, who are in the same classroom. We further found that 41% of the variation in maternal and teacher ratings is explained by a common factor. The heritability of this common factor is 78%. The heritabilities of the rater specific factors of mothers and teachers are 76% and 39%, respectively. Because Attention Problems that are persistent over situations may indicate more serious behavior problems than context dependent Attention Problems, we believe that gene finding strategies should focus on this common phenotype.

INTRODUCTION

Assessing whether a young child has Attention Problems is difficult. In order to quantify and characterize Attention Problems, researchers and clinicians often have to rely on ratings of parents and teachers. These ratings may be influenced by the rater's personal values or perspective, and by the unique settings in which the rater and child co-exist. Agreement between raters suggests that some aspects of the rated behavior can be reliably assessed regardless of rater or situation. Three different explanations exist for rater disagreement. First, different raters may assess unique aspects of the behavior, which are situation or context dependent. For example, a child's inability to concentrate or to sit still may be obvious in the classroom setting, but less evident in other settings, where sustained attention is less important (e.g., at home). Second, parents and teachers may have different perspectives to the child's behavior. The perspectives may differ for a variety of reasons; teachers are not biologically related to the children, and they are exposed to the behavior of many children of the same age. Third, raters may show rater bias, i.e., their ratings are influenced by their own personal norms and values.

In studies of the teacher and parent ratings of the same children, the agreement between these informants is modest. Ende van der & Verhulst (2005) reported parent-teacher correlations on AP in the range of .29 to .41 in a sample of Dutch boys and girls in two different age-groups. Achenbach & Rescorla (2001) found a correlation of .44 between parent and teacher ratings on Attention Problems. The FinnTwin12 study reported higher correlations for ratings on inattention, which pertained to the same setting, e.g., ratings of teachers and classmates, than the correlations for ratings which pertained to different settings, e.g., ratings of teachers and parents, or of classmates and parents (Pulkkinen et al., 1999). The latter findings imply that behavior is in part context dependent.

Previous twin studies supported the hypothesis that mothers and teachers have different perspectives on children's levels of hyperactivity and Attention Problems (Nadder & Silberg, 2001) and ADHD (Martin et al., 2002; Thapar et al., 2000). To determine how much of the variation in parent and teacher ratings is due to rating similar versus situation specific components of behavior, some investigators employed bivariate model fitting analyses, which revealed that maternal and teacher ratings on hyperactivity partly reflect a common latent phenotype (Martin et al., 2002; Simonoff et al., 1998). In addition to this common phenotype, maternal ratings reflected rater contrast effects, while teacher ratings reflected aspects of the children's behavior, which did not influence maternal ratings (Simonoff et al., 1998).

It has been shown convincingly that variation in children's inattentive and hyperactive/impulsive behavior is attributable to both genetic and environmental factors. Heritability estimates of parent ratings on AP and/or Hyperactivity (HI) usually vary between 50 and 80% (Hudziak et al., 2005; Rietveld et al., 2003b; Martin et al., 2002; Hudziak et al., 2000). The heritability estimates of teacher ratings on AP and/or HI tend to be lower than those of parent ratings, and usually fall in the range of 40 to 70% (Vierikko et al., 2004; Kuntsi & Stevenson, 2001; Thapar et al., 2000; Eaves et al., 1997; Sherman et al., 1997). The study of Martin et al. (2002), in which the number of ADHD-symptoms was established in 5 to 16-year-old children, is the only one, in which heritability estimates were slightly lower in parent ratings (74%) than in teacher ratings (80%).

An interesting finding is that parent and teacher ratings differ not only in the size of the heritability estimate, but also in the etiology of the sources of individual differences. Parent ratings on ADHD are often characterized by non-additive genetic effects (Martin et al., 2002), or contrast effects (Eaves et al., 1997; Kuntsi & Stevenson, 2001), while teacher ratings are not. These differences are evident in the corrrelations of the parent ratings of ADHD, which are often very low in DZ twins (Simonoff et al., 1998; Eaves et al., 1997), while teacher ratings do not show these low correlations. Low DZ correlations can be

explained either by the presence of non-additive genetic effects (Lynch & Walsh, 1998), or by contrast effects (Eaves, 1976). These two phenomena both predict low DZ correlations, but the presence of a contrast effect also predicts different variances in MZ and DZ twins. Theoretically the two can thus be distinguished, although Rietveld et al. (2003a) have shown that the statistical power of the classical twin study to do so is low. A further complication is that, given only parent ratings, one cannot distinguish between a contrast effect on the phenotypic level (sibling interaction), and a contrast effect on the observed level (rater bias). Therefore, Simonoff et al. (1998) simultaneously analyzed parent and teacher ratings on childhood hyperactivity. They found that the contrast effect in parent ratings was due to rater bias, not to sibling interaction.

With respect to teacher ratings, it is often the case that correlations are higher in children rated by the same teacher than correlations in children rated by different teachers (Saudino et al., 2005; Vierikko et al., 2004; Towers et al., 2000; Simonoff et al., 1998; but not in Sherman et al., 1997). Higher correlations in children rated by the same teacher than in children rated by different teachers, suggest that teacher rater bias plays a role. Simonoff et al. (1998) developed two different models to explore this finding. One model was based on the assumption that teachers have difficulty distinguishing the two children ("twin confusion model"). The other model was based on the assumption that ratings by the same teacher are correlated, because a) raters have their own subjective perspective on which behaviors are (in)appropriate, or b) raters themselves influence the behavior of the child, as a function of his/her (i.e., the rater's) own personality characteristics ("correlated errors model"). However, in their sample of 1044 twin pairs, Simonoff et al. were not able to differentiate between the twin confusion and the correlated errors model. A complicating factor in analyzing behavioral ratings of the same versus different teachers is that classroom separation may not be a random process. In Dutch twins, separation is somewhat more likely when children score high on externalizing problems at age three (Leeuwen van et al., 2005).

In the present paper, we will examine the contribution of genetic and non-genetic factors to individual differences in Attention Problems (AP). By analyzing maternal and teacher ratings, we estimate the extent to which the agreement between maternal and teacher reports on childhood AP is caused by the same genetic and/or environmental factors being expressed in different surroundings (e.g., the classroom versus the home). Given the size and nature of our twin sample, we are also able to test the contribution of teacher rater bias, as approximately half of our sample is placed into same and half into different classrooms. Although maternal ratings may also be prone to rater bias, we can not directly test for this because twins are always rated by the same mother.

METHODS

SUBJECTS

This study is part of an ongoing longitudinal twin study in the Netherlands. The subjects were all registered at birth with the Netherlands Twin Register (Boomsma et al., 2002). For the present study, we analyzed data of a sample of Dutch twins, whose mothers and teachers reported on their behavior, when they were seven years old. The twins were born between 1992 and 1996. Maternal ratings were available for 2,310 complete twin-pairs and 8 incomplete twin-pairs and teacher ratings were available for 2,276 complete twin-pairs and 281 incomplete twin-pairs. In 86% of the twins, ratings were available for both mothers and teachers, in 5% only from mothers, and in 9% only from teachers. Furthermore, about 53% of the twins were in the same classroom, while 36% of the twins were in different classrooms. Of the remaining 11% of the sample, it was unknown whether they were in the same or in different classrooms, mainly due to the fact that a teacher questionnaire was returned

for only one of the children. Twin-pairs for whom it was unknown whether the two members of the pair were rated by the same or by different teachers were excluded from the analyses.

Zygosity diagnosis was based on DNA in 123 same-sex twin pairs. In the remaining same-sex pairs, zygosity was assessed with the use of a 10-item questionnaire. This procedure allows an accurate determination of zygosity of nearly 95%. It is described in more detail in Rietveld et al. (2000). The pairs of whom zygosity status could not be determined (N=31 pairs) were excluded from the analyses. The number of twin pairs, by sex, zygosity, and informant are presented in Table 1.

PROCEDURE

A survey, including the CBCL/4-18, was mailed to the mothers of the twins when the twins were seven years old. Mothers, who did not return the forms within two months, received a reminder. Where financially possible, persistent non-responders were contacted by phone four months after the initial mailing. This procedure resulted in a 66% participation rate. Rietveld et al. (2004) showed that non-participation at age 7 is positively related to the twin's overactive behavior at age 3. However, the difference in overactive scores at age 3 between mothers who do respond (mean=2.76), and mothers who do not respond (mean=2.86) at age 7 is small. Once the parent's permission was procured to approach the teacher, a Teacher Report Form (TRF) was sent to the teacher. After two months, a reminder was sent to the nonresponding teachers. The participation rate of the teachers was 78% (Leeuwen van et al., 2005). The number of teacher ratings is greater than the number of maternal ratings due to different time schedules for the data entry.

Table 1. Number of twin pairs (complete/incomplete)

	SAME TEACHER		DIFFERENT TEACHER	
	M	T	M	T
MZM	209/0	236/0	126/2	140/1
DZM	184/1	194/0	153/0	166/2
MZF	247/1	260/1	162/0	177/2
DZF	182/0	196/1	132/1	147/0
DOS	399/2	433/4	252/1	270/3
Zygosity unknown	2/0	15/0	1/0	13/0
Total	1223/4	1334/4	826/4	913/8

M=mother; T=teacher; MZM=monozygotic male; DZM=dizygotic male,
MZF=monozygotic female; DZF=dizygotic female; DOS=opposite sex twins

MEASURES

The Child Behavior Checklist (CBCL/4-18) (Achenbach, 1991a) contains 120 items that measure problem behavior. The items are rated on a 3-point scale ranging from "not true", "somewhat or sometimes true", to "very true or often true". In the present paper, we report on the Attention Problem scale (11 items). The two week test-retest correlation and the internal consistency in this age group are .83 and .67, respectively (Verhulst et al., 1996).

The Teacher Report Form (TRF) (Achenbach, 1991b) contains 120 items that measure problem behavior with the same three response categories as the CBCL. The Attention Problems scale contains 20 items. The six week test-retest correlation is .83. The internal consistency coeffients are .90 and .92 in boys and girls, respectively (Verhulst et al., 1997). Ten items overlap between the AP scales of the CBCL and the TRF.

Statistical Analyses

Both the TRF and CBCL data show high skewness (1.56, and 1.43, respectively), and high kurtosis (2.23, and 2.40, respectively). Derks, Dolan & Boomsma (2004a) showed that bias in parameter estimates due to non-normality of the data may be avoided by using categorical data analysis. In this approach, a liability threshold model is applied to the ordinal scores (Lynch & Walsh, 1998). It is assumed that a person is "unaffected", if his or her liability is below a certain threshold, and that he or she is "affected", if his or her liability is above this threshold. In the present paper, the CBCL and TRF scores were recoded in such a way that three thresholds divide the latent liability distribution into four categories. The thresholds are chosen in such a way that the prevalences are more or less similar in each of the four categories.

In order to test whether the prevalences of Attention Problems vary by sex, or by same and different teacher, we compared the fit of a model in which the thresholds are equated with the fit of a model in which the thresholds are allowed to be different. All statistical analyses were performed in Mx (Neale et al., 2003). The type-I error rate of all statistical tests was set at .01 (rather than .05) to accommodate multiple testing.

Genetic Modeling

Genetic analyses were performed in a multi-group design of MZM (monozygotic males), DZM (dizygotic males), MZF (monozygotic females), DZF (dizygotic females), and DOS (opposite sex twins). In addition, the twins were divided into a same teacher group and a different teacher group. This resulted in a ten-group analysis.

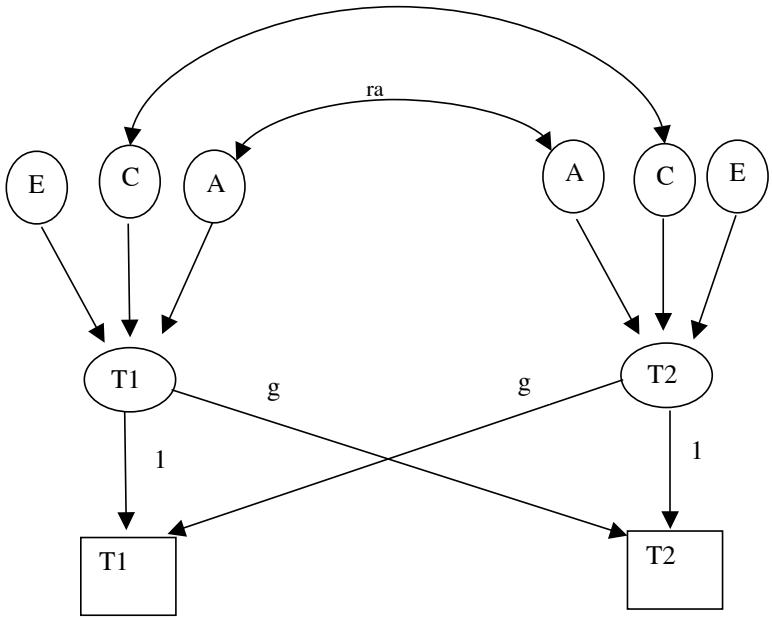
Univariate genetic models were fitted to maternal ratings on AP. We analyzed the data of children in the same and different classrooms separately for three reasons. First, Leeuwen van et al. (2005) showed that in Dutch twins, separation is somewhat more likely, when children score highly on externalizing problems at age three. If these mean differences persist to age 7, combining data from children in same versus different classrooms may give biased estimates of the correlations. Second, Simonoff et al. (1998) reported a slightly higher heritability in maternal ratings for children, who are in the same classroom than for children, who are in different classrooms. Third, it could be the case that children who have the same teacher become more similar, because of their greater environmental sharing at school.

A fully saturated model, in which all correlations and thresholds were freely estimated, was fitted to the ordinal data. Next, we examined whether the thresholds differed between MZ and DZ twins. Because contrast effects cause different variances in MZ and DZ twins, and therefore lead to different prevalences of Attention Problems among these groups, contrast effects were only included if the thresholds of MZ and DZ twins were different. Third, a model that includes additive genetic effects (A), shared environmental (C), or dominant genetic effects (D), and nonshared environmental effects (E) was fitted to the data. It should be noted that the effects of C and D cannot be modeled simultaneously, as they are not both identified. If the correlations in MZ twins were more than twice the correlations in DZ twins, D was included in the model. If the correlations in MZ twins were less than twice the DZ correlations, C was included. Finally, a series of more parsimonious models were fitted: a) variance components in the best fitting model were constrained to be equal in boys and girls; b) variance components A, and C or D were constrained at zero; c) the variance components were constrained to be the same for children in the same classroom and children in different classrooms. The fit of the more parsimonious models were compared with the fit of the full model by means of the likelihood ratio test.

The univariate models that were fitted to the teacher ratings on AP were based on the models that were presented in Simonoff et al. (1998). Similarly to the model fitting of the maternal data, a fully saturated model was fitted to the data. Next, two different genetic models were fitted. In the "twin confusion" model (see Figure 1), the higher twin correla-

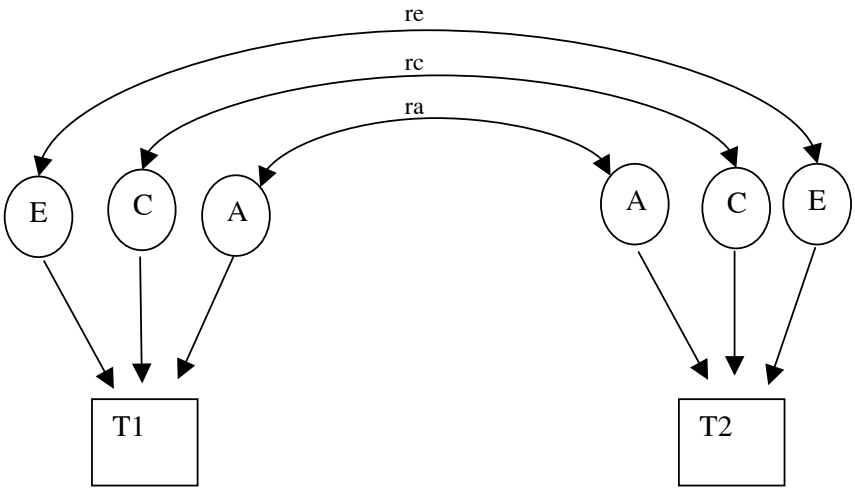
tions in pairs rated by same teachers are explained by the fact that teachers may not always distinguish between the two individuals in a twin pair. The confusion paths are allowed to differ according to zygosity, because we expect more confusion in MZ twins than in DZ twins. Furthermore, the confusions paths are assumed to be absent when children are rated by different teachers or when the individuals in a twin pair are of opposite sex. The second model, the "correlated errors" model (see Figure 2), specifies that teachers bring their own influences into their ratings of behavior either because they have their own subjective perspective, or because they influence the behavior of the child, as a function of rater bias (i.e., the rater's own personality characteristics) (Simonoff et al., 1998). When all twin-pairs are rated by the same informant, rater bias is shared between two members of a twin-pair and is therefore confounded with shared environmental influences. Because we have access to data from twins who are rated by the same teacher and from twins who are rated by different teachers, we are able to distinguish between true shared environmental influences and rater bias. In the correlated error model, the nonshared environmental component is allowed to correlate in children that are rated by the same teacher. If this correlation is significantly greater than zero, this may be evidence of teacher rater bias. It should be noted that the term "correlated error" is too restrictive, because the children rated by same teachers may actually behave more alike, because of certain characteristics of the teacher and/or classmates (e.g., teaching styles, social interactions in the group), or by the fact that classroom separation depends on the level of problem behavior before separation. In these cases, the higher correlation would not be caused by error. However, to avoid confusion, we choose to retain the original name of the model.

Figure 1. Twin confusion model for attention problem scores of children rated by thesame versus different teachers



Note: Latent factors are represented as circles, observed variables are represented as squares. A=additive genetic effects; C=shared environmental effects; E=non-shared environmental effects; g=twin confusion path (the loading of twin on his/her cotwins attention problem score); T1(circle)=latent AP score twin 1; T1(square)=observed score twin 1; T2(circle)=latent AP score twin 2; T2(square)=observed score twin 2; ra=1(MZ) or .5(DZ); rc=1(MZ and DZ). The loading g is allowed to vary as a function of zygosity. In opposite sex twins, and in twin-pairs in which both members of the pair are rated by different teachers, the loading g is constrained at zero. The total variance of the latent factor is constrained at 1.

Figure 2. Correlated error model for attention problem scores of children rated by the same versus different teachers



Note: Latent factors are represented as circles, observed variables are represented as squares. A=additive genetic effects; C=shared environmental effects; E=non-shared environmental effects; e=); T1(square)=observed score twin 1; T2(square)=observed score twin 2; ra=1(MZ) or .5(DZ); rc=1(MZ and DZ); re=correlated error path which is constrained at zero in twin-pairs who are rated by different teachers, and is freely estimated in twin-pairs who are rated by the same teacher.

Finally, we fitted a bivariate psychometric model to maternal and teacher data. In the psychometric model (Hewitt et al., 1992), the ratings of different informants are allowed to be influenced by a common behavioral view and shared understanding of the behavioral descriptions, and also by unique aspects of their child's behavior. In the bivariate model, we included common factors that influence both maternal and teacher ratings, specific maternal factors that influence maternal ratings only, and specific teacher factors that influence teacher ratings only. Based on the results of the univariate analyses, we identified the most appropriate bivariate model. If rater disagreement is the result of rater bias, the twin correlations of the rater specific factors would not depend on zygosity, and the rater specific variance would be explained by shared environmental influences. In contrast, when the rater disagreement is the result of each rater assessing unique aspects of the child's behavior, and given that the trait is heritable, we would expect to find genetic influences on the rater specific variance.

RESULTS

Prevalence of AP

Mean scores, standard deviations, and thresholds for maternal and teacher reports on AP are summarized in Table 2. Differences in the distribution of AP were examined by equating the thresholds in Mx. Boys obtained higher AP scores than girls ($\chi^2(48)=165.14$, $p<.001$; $\chi^2(48)=223.14$, $p<.001$, for maternal and teacher ratings respectively). Maternal and teacher AP scores did not differ between children in different classrooms and children in the same classroom ($\chi^2(36)=54.31$, $p=.03$; $\chi^2(36)=34.20$, $p=.55$, for maternal and teacher ratings, respectively).

Table 2. Mean and standard deviations (SD) of raw scores, and thresholds of maternal and teacher ratings on Attention Problems in seven year old boys and girls

	Mother, same classroom		Mother, different classrooms		Teacher, same classroom		Teacher, different classrooms	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Mean	2.86	2.23	3.37	2.44	6.17	3.73	7.37	4.31
SD	2.64	2.45	3.00	2.75	6.41	5.08	7.05	5.24
T1	-.91	-.55	-.98	-.60	.04	.57	-.11	.43
T2	.06	.40	-.05	.36	.77	1.33	.63	1.12
T3	1.00	1.35	.77	1.14	1.50	1.89	1.36	1.87

T1=threshold 1; T2=threshold 2; T3=threshold 3

Twin correlations

Polychoric twin correlations were estimated for each sex-by-zygosity group in Mx. The maternal and teacher cross-twin correlations represent the agreement between the twins within each rater. The within-twin cross-rater correlations represent the agreement between the raters within the same child. Finally, the cross-twin cross-rater correlations represent the agreement between raters between the two members of a twin pair. One example of the latter is the correlation between the maternal rating of the first-born twin and the teacher rating of the second born twin.

The correlations of the maternal and teacher AP-scores are shown in Table 3. Because only 10 items overlap between the maternal and teacher AP-scales, we also calculated the correlations on the basis of the 10 overlapping items (see Table 4). The correlations of the overlapping items are no higher than the correlations of the original AP-scales. To facilitate the comparison of the results of the genetic analyses with those of other studies using the TRF, we chose to perform the statistical analyses on the original scales.

Regardless of informant, MZ twin correlations are higher than DZ twin correlations, which suggests the presence of genetic influences. The maternal cross-twin correlations are more than twice as high in MZ twins as in DZ twins, which is suggestive of genetic dominance. Therefore, we fitted an ADE model to the maternal ratings. Because the teacher cross-twin correlations are less than twice as high in MZ as in DZ twins, we fitted an ACE model to the teacher ratings. The cross-twin cross-rater correlations, which represent the common part of the maternal and teacher ratings, are much higher in MZ twins than in DZ twins. We would therefore expect an ADE model to provide the best fit to the common part of the bivariate model.

Statistical tests showed that maternal correlations did not differ among twin pairs in the same classroom versus different classrooms ($\chi^2(6)=15.52$, $p=.02$). In contrast, teacher correlations were higher in children rated by the same teacher than in children rated by different teachers ($\chi^2(6)=40.89$, $p<.001$).

Table 3. Polychoric Correlations of the maternal and teacher ratings on AP

		Maternal cross-twin	Teacher cross-twin	Within twin Cross-rater	Cross-twin Cross-rater
MZM	ST	.77	.81	.48	.40
	DT	.79	.56	.51	.38
DZM	ST	.39	.49	.32	-.04
	DT	.25	.22	.51	.02
MZF	ST	.80	.82	.43	.37

ST=Same Teacher; DT=Different Teacher; MZM=monozygotic male; DZM=dizygotic male, MZF=monozygotic female; DZF=dizygotic female; DOS=opposite sex twins

Table 3. Polychoric Correlations of the maternal and teacher ratings on AP *(Continued)*

		Maternal cross-twin	Teacher cross-twin	Within twin Cross-rater	Cross-twin Cross-rater
	DT	.69	.51	.37	.25
DZF	ST	.35	.48	.36	.00
	DT	.35	.39	.39	.13
DOS	ST	.16	.49	.34	-.05
	DT	.20	.21	.42	.03

ST=Same Teacher; DT=Different Teacher; MZM=monozygotic male; DZM=dizygotic male, MZF=monozygotic female; DZF=dizygotic female; DOS=opposite sex twins

Table 4. Polychoric Correlations of the 10 overlapping items of the maternal (M) and teacher (T) AP scales

		Maternal cross-twin	Teacher cross-twin	Within twin Cross-rater	Cross-twin Cross-rater
MZM	ST	.77	.80	.47	.43
	DT	.79	.43	.47	.34
DZM	ST	.39	.44	.29	-.07
	DT	.24	.31	.48	.08
MZF	ST	.80	.83	.38	.32
	DT	.69	.44	.32	.17
DZF	ST	.35	.44	.29	.02
	DT	.36	.34	.28	.13
DOS	ST	.32	.36	.35	-.01
	DT	.19	.24	.35	.00

ST=Same Teacher; DT=Different Teacher; MZM=monozygotic male; DZM=dizygotic male, MZF=monozygotic female; DZF=dizygotic female; DOS=opposite sex twins

Univariate genetic model fitting analyses of maternal ratings

In maternal ratings on AP, the thresholds did not differ between MZ and DZ twins ($\chi^2(24)=30.10$, p=.18). Therefore, rater contrast effects were not included in the genetic model. The results of the genetic model fitting are summarized in Table 5. The best-fitting model is printed in bold. Briefly, the univariate genetic analyses showed significant influences of A, D, and E. The estimates of A, D, and E did not depend on sex, and did not differ among twins in the same classroom and twins in different classrooms. The relative influences of A, D, and E in the best fitting model were 44%, 33%, and 23%, respectively. Compared to a saturated model, the fit of this model was good ($\chi^2(10)=18.30$, p=.932).

Table 5. Univariate model fitting of maternal Attention Problem ratings in 7-year-old children

MODEL	PARAMETERS	-2 LL	WITH MODEL	ΔDF	χ^2	P
1. Fully saturated model	84	10201.14	-	-	-	-
2. ADE model, boys girls, same different teacher	80	10210.80	1	4	9.66	.047
3. ACE model, boys girls, same different teacher	80	10224.04	1	4	22.90	.000
4. ADE model, boys = girls, same different teacher	76	10216.21	2	4	5.41	.248
5. ADE model, boys = girls, same = different teacher	**74**	**10219.44**	**4**	**2**	**3.23**	**.199**
6. AE model, boys = girls, same = different teacher	73	10226.86	5	1	7.42	.006

A = additive genetic effects, C=shared environmental effects, D=dominant genetic effects, E=non-shared environmental effects, -2LL=-2 log likelihood, df=degrees of freedom
boys=girls: equating the non-standardized parameters of boys and girls
same=different teacher: equating the non-standardized parameters of same and different teachers

UNIVARIATE GENETIC MODEL FITTING ANALYSES OF TEACHER RATINGS

The results of the model fitting analyses on teacher ratings are shown in Table 6. An ACE model that allowed for different influences of A, C, and E in same and different teachers provided a good fit to the data. However, the more parsimonious "correlated errors" model also provided a good fit. The "twin confusion" model did not fit well. In the correlated errors model, the relative influences of genes and environment did not differ between boys and girls, and the influence of the shared environment was not significant. The heritability of teacher ratings on AP was 55% and the nonshared environment explained 45% of the variation. The nonshared environment correlated .54 when children were rated by the same teacher. This correlation was significantly greater than zero. Compared to a saturated model, the fit of the correlated errors model was good ($\chi^2(10)=6.79$, p=.745).

Table 6. Univariate model fitting of teacher Attention Problem ratings in 7-year-old children

MODEL	PARAMETERS	-2 LL	WITH MODEL	ΔDF	$\Delta\chi^2$	P
1. Fully saturated model	84	8683.69	-	-	-	-
2. ACE model, boys girls, same different teacher	80	8685.37	1	4	1.68	.794
3. ACE, Correlated error model, boys girls	78	8688.55	1	6	4.86	.562
4. ACE, Twin confusion model, boys girls	80	8694.12	1	4	10.43	.034
5. ACE, Correlated error model, boys = girls	75	8690.48	3	3	1.93	.587
6. AE, Correlated error model, boys = girls	**74**	**8690.48**	**5**	**1**	**.00**	**-**
7. AE, boys = girls, correlated error dropped	73	8726.41	6	1	35.93	.000

A= additive genetic effects, C=shared environmental effects, D=dominant genetic effects, E=non-shared environmental effects, -2LL=-2 log likelihood, df=degrees of freedom
boys=girls: equating the non-standardized parameters of boys and girls
same=different teacher: equating the non-standardized parameters of same and different teachers

BIVARIATE GENETIC MODEL FITTING ANALYSES OF MATERNAL AND TEACHER RATINGS

Based on the results of the univariate genetic analyses, we fitted a bivariate model that included a common part consisting of the factors Ac, Dc, and Ec, a unique maternal part consisting of the factors Am, Dm, and Em, and a unique teacher part consisting of the factors At and Et. A correlated error was only included in the unique teacher part of the bivariate model. The bivariate model fitting results are summarized in Table 7. In the best fitting bivariate model, 41% of the variation in maternal and teacher ratings on AP was explained by a common factor. This common factor was decomposed into a dominant genetic factor, which explained 32% of the total variation, and a non-shared environmental factor, which explained 9% of the total variation. The heritability of the common factor is 78% (this can be calculated as the amount of variation explained by genetic factors divided

by the total variance .32/.41=78%). Variation in maternal ratings was further explained by Am (45%), and Em (14%). Variation in teacher ratings was explained by At (23%), and Et (36%), and a correlated error of .77 in same teacher ratings. Compared to a saturated model, the fit of the bivariate model was not very good ($\chi^2(66)=103.09$, p=.002). However, it is known that in bivariate analyses, the power to detect very small differences is high. Therefore, we calculated the residuals of the expected covariance matrices of the ADE model and the expected covariance matrices under the saturated model. Expectation of these residuals showed that the misfit was mainly different cross-rater cross-twin correlations in the monozygotic male group rated by the same teacher (i.e., the correlation of teacher-firstborn with mother-second born is not equal to the correlation of mother-firstborn with teacher-second born. Because there is no theoretical reason why this correlation would depend on birth-order, we accepted the ADE model as the best-fitting model. This model is shown in Figure 3, including the estimated factor loadings. As an illustration of Figure 3, we will show how the heritability of the common factor can be derived based on the factor loadings. The total variance of the common factor is $.30^2 + .57^2 + .00^2 = 0.41$. The variance explained by the genetic factor is $.57^2 = .32$. Therefore, the proportion of the variance explained by genetic factors = .32/.41=.78, and the heritability of the common factor is 78%. Figure 4 gives an overview of the genetic and environmental influences on the common and rater-specific parts of the model.

DISCUSSION

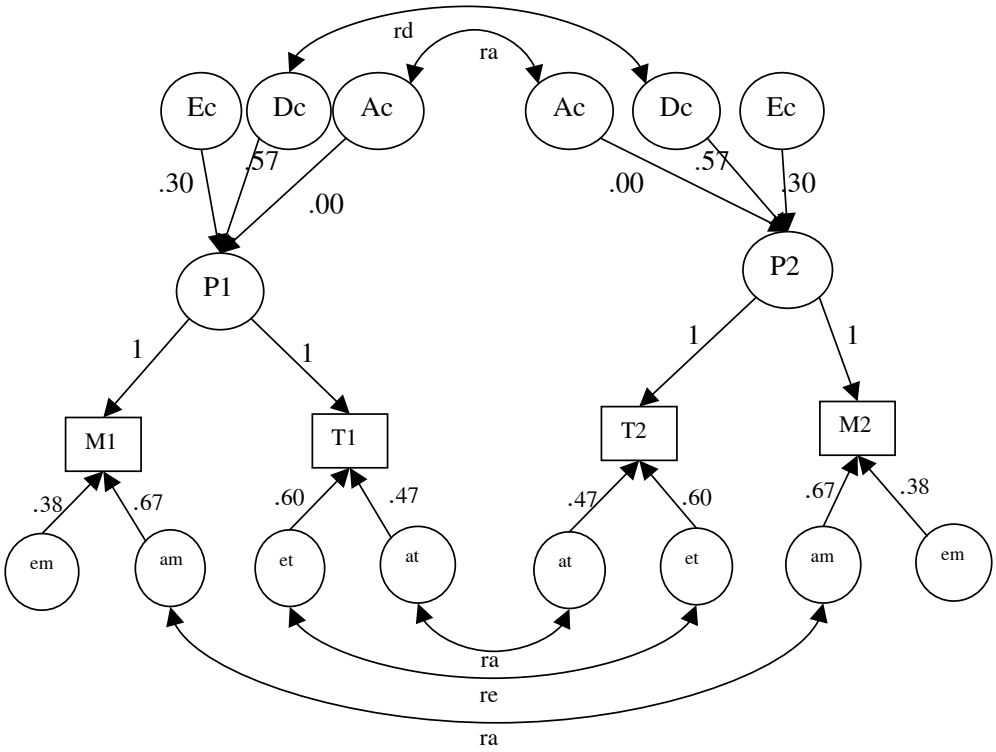
The purpose of this study was to examine the genetic and environmental contributions to the variation in maternal and teacher ratings on Attention Problems in children, and to the covariation between these ratings. In the univariate genetic analyses, the heritability estimate was higher in maternal ratings (77%) than in teacher ratings (54%), which agrees with previous findings (Vierikko et al., 2004; Kuntsi & Stevenson, 2001; Thapar et al., 2000; Simonoff et al., 1998; Eaves et al., 1997; and Sherman et al., 1997). A more thorough investigation of the correlations, however, revealed that the correlation in the maternal data was similar to the correlation in the 'same teacher' data, and that both of these were higher than the correlation in the 'different teacher' data. These data therefore support the inference that the lower heritability in teacher ratings is due to combining data from same and different teachers. This is consistent with the findings of Martin et al. (2002), who observed similar heritabilities in parent and teacher ratings in a sample consisting for 91% of children rated by same teachers. In contrast, Vierikko et al. (2004) also conducted genetic analyses on twin-pairs, in which both members were rated by the same teacher, and reported a lower heritability in teacher (49-55%) than parent (78-81%) ratings.

Table 7. Bivariate model fitting of maternal and teacher Attention Problem ratings in 7-year-old children

MODEL	PARAMETERS	-2 LL	WITH MODEL	ΔDF	$\Delta\chi^2$	P
1. Fully saturated model	216	18362.96	-	-	-	-
2. ADE model, boys girls	160	18460.51	1	56	97.55	.000
3. ADE model, boys = girls	152	18466.05	2	8	5.54	.699
4. ADE model, boys = girls, rater-specific D dropped	**150**	**18466.05**	**3**	**2**	**0.00**	**-**

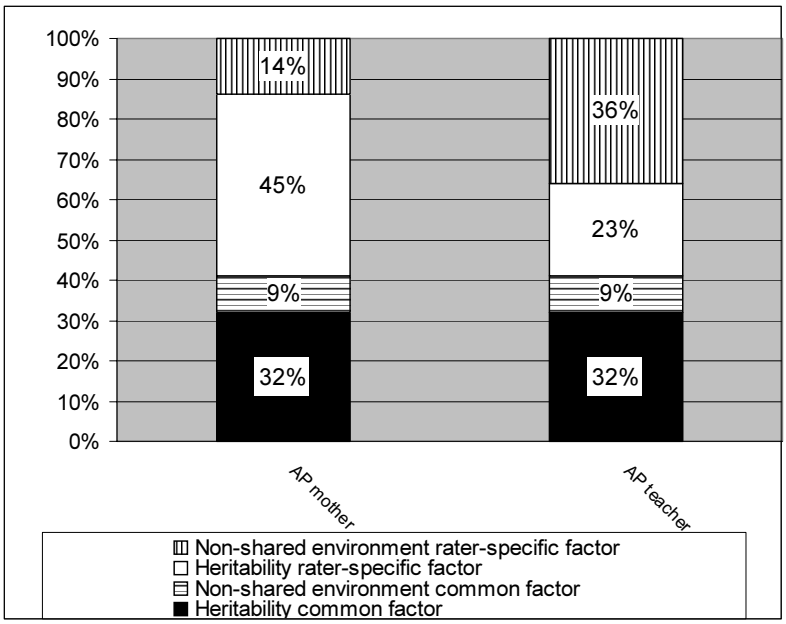
A=additive genetic effects, C=shared environmental effects, D=dominant genetic effects, E=non-shared environmental effects, -2LL=-2 log likelihood, df=degrees of freedom; boys=girls: equating the non-standardized parameters of boys and girls

Figure 3. Bivariate model for maternal and teacher ratings



Note: P1=phenotype twin 1; P2=phenotype twin 2; M1=Maternal rating of twin 1; M2=Maternal rating of twin 2; T1= Teacher rating of twin 1; T2=Teacher rating of twin 2. Ac, Dc, and Ec are the common additive genetic, dominant genetic, and non-shared environmental effects; em and am are the unique maternal additive genetic, dominant genetic, and non-shared environmental effects; et and at are the unique teacher additive genetic, dominant genetic, and non-shared environmental effects.ra=1(MZ) or .5(DZ); rd=1(MZ) or .25(DZ); re is the correlated error path and is estimated at .77. It is constrained to be equal inm MZ and DZ twins, and is assumed to be absent in children rated by different teachers. The paths from the latent phenotypes P1 and P2 to the maternal and teacher ratings are constrained at 1.

Figure 4. Graphical representation of the influences of genes and non-shared environment on Attention Problems in seven year old twins



In summary, the pattern is somewhat inconsistent, but the present results suggest that the higher heritabilities in parental ratings than teacher ratings can be explained by the fact that twins are always rated by the same parent, but in about half of the cases by different teachers.

Previously, it was shown that the higher twin correlations in children rated by the same teacher than in children rated by different teachers are associated with a higher heritability of problem behavior (Saudino et al., 2005; Simonoff et al., 1998). Simonoff et al. (1998) compared the fit of two distinct theoretical models to explain this finding, but both models fit equally well. In the present study, we were able to differentiate between these models; the correlated error model provided a better fit to the data than the twin confusion model. The correlated error may be caused by rater bias, reflecting the fact that raters have their own specific perspective on which behaviors are (in)appropriate. An alternative explanation is that the correlated error reflects true qualities of the children's behavior, which are elicited by the exposure to a particular rater (Simonoff et al., 1998). For example, different teachers may elicit different behaviors from children.

These two alternative explanations have different implications for the interpretation of the high correlations in children who are rated by the same informant. If it is true that the correlations are higher because of rater specific views, this implies that the phenotypic correlations in both maternal and same teacher ratings are overestimated, and that the influence of non-shared environmental factors is underestimated. Alternatively, if the higher correlations are the result of the fact that children behave more similarly when confronted with the same person, this suggest that the behavior of children depends on the person, with whom they interact. In this case, the lower correlation in twins rated by different informants is the result of an increase in the non-shared environmental variance, and the high correlations, when twins are rated by the same informant, reflect the true phenotypic similarity of children interacting with the same person. The second possibility may explain the higher correlation between paternal and maternal ratings than the correlation between parent and teacher ratings. The parents usually observe the children in interaction with the

other parent, but not in interaction with the teacher. Consequently, correlations should be lower in parents, who are divorced, than in parents, who live together. With the available data, we cannot decide whether the high correlation in twins rated by the same informant are caused by rater specific views, or by the influence of the informant on the child's behavior.

As in previous studies on Attention Problems (Vierikko et al., 2004; Ende van der & Verhulst, 2005), the correlations between maternal and teacher ratings were moderate. We ruled out the possibility that rater disagreement is the result of non-overlapping items of the AP-scales of parents and teachers by showing that the correlations of the overlapping items are not higher than the correlations of the original scales.

The bivariate model fitting analyses showed that slightly less than half of the variation in maternal and teacher ratings is explained by common aspects of the child's behavior while the remaining variation is explained by rater or setting specific aspects. The finding of genetic influences on the rater specific variance shows that the diagreement between parents and teachers is not solely due to rater bias. Both raters assess unique aspects of the child's behavior. The common aspects, which are highly genetic, reflect the part of the phenotype that is stable across settings and raters. The genetic variation of the common factor was completely explained by dominant genetic effects. This is in agreement with the low cross-twin cross-rater correlations reported in DZ twins. The large dominant genetic influences were surprising because these were not found in the univariate analyses of the teacher ratings. How can we explain the low cross-twin cross-rater correlations in DZ twins? The fact that low cross-rater cross-twin correlations (-.12-.21) were also reported by Simonoff et al. (1998) in their study on hyperactivity suggests that these are not the result of artifacts in our data collection. The presence of sibling interaction is not likely either as these effects should also be found in the univariate analyses of maternal or teacher ratings. Rater bias does not seem to play a role, because it is hard to envisage that high teacher ratings on AP in twin 1 would lead to low maternal ratings on AP in twin 2. The only explanation that we can offer here is that variation in maternal and teacher rating is influenced by a correlated error, which increases the correlation in MZ and DZ twins, and mimics the effect of shared environmental influences, and by dominant genetic effects. These effects might cancel each other out in the univariate analyses (which would suggest that the dominance effect reported for maternal ratings is underestimated). In the common factor of the bivariate analyses (i.e., the factor that influences both maternal and teacher ratings), correlated errors are absent, and the presence of the dominance genetic effects is evident. Some support for this explanation is provided by the different teacher correlations. In boys, the different teacher correlations show a pattern that is in agreement with the presence of genetic dominance while the same teacher correlations do not. However, in girls, the pattern of the different teacher correlations in DZ girls is not suggestive of genetic dominance. Future studies should reveal further insight regarding the low cross-twin cross-rater correlations.

The significant influence of genes on the rater-specific aspects is consistent with Martin et al. (2002), who found that variation in maternal and teacher ratings on hyperactivity is partly influenced by different genes. It implies that disagreement between parents and teachers is not merely due to rater bias. This finding is consistent with the results of Bartels et al. (2004) and Derks et al. (2004b), who found that mothers and fathers assess unique aspects of the child's behavior, although most variation in these ratings is explained by common aspects. Apparently, mothers and teachers both rate meaningful, but partly different, aspects of children's behavior.

The fact that the prevalence of ADHD was similar in MZ and DZ twins, is supportive of an absence of rater contrast or sibling interaction in maternal ratings. In the literature on AP and HI, contradictory findings are reported with respect to the presence of contrast

effects in parental ratings. Significant contrast effects on AP and/or HI have been reported in some studies (Simonoff et al., 1998; Kuntsi & Stevenson, 2001; Vierikko et al., 2004; Eaves et al., 1997; Eaves et al., 2000), but not in others (Kuntsi et al., 2005; Martin et al., 2002, Thapar et al., 2000, Towers et al., 2000, and Hudziak et al., 2000). In teacher ratings, contrast effects are absent (Simonoff et al., 1998; Kuntsi & Stevenson, 2001; Vierikko et al., 2004; Eaves et al., 1997). Plomin (1982) suggest that contrast effects are more likely when the items refer to global descriptions of behavior rather than to specific descriptions of behavior. This was confirmed by Saudino et al. (2004), who report a tendency for contrast effects to be more pervasive when global ratings were required. The lack of contrast effects in the current study shows that the items of the CBCL are specific enough to prevent parents from comparing the behavior of the twins.

The results should be interpreted with the following points kept in mind. First, the CBCL and TRF do not assess the presence of DSM symptoms. The CBCL-AP scale does predict the presence of DSM-IV ADHD (Hudziak et al., 2004), but whether this is so for the teacher form is unknown. There are a number of reasons that DSM interviews of teachers are rarely employed, including time burden and expense. However, perhaps the most important reason is the lack of an empirically validated DSM-IV teacher data base. Thus, although teacher reports on ADHD are commonly used, there is little known about the validity of these reports. The Netherlands Twin Register is currently collecting data on school performance, and in future studies we will address the question whether high teacher AP-scores are more predictive for problems related to school performance than high parental AP-scores. Second, the results in this study are based on analysis of Attention Problems rated by parents and teacher. It is unclear whether the results generalize to hyperactivity. However, the Attention Problem scales do include some items on hyperactivity (e.g., Can't sit still, restless, or hyperactive), and a review of epidemiological genetic studies shows that the heritabilities of Attention Problems and Hyperactivity are similar (Derks et al., in press).

Higher correlations are found in children rated by the same informant than in children rated by different informants. At this point, it is unclear whether the higher correlation based on ratings from the same informant overestimate the true phenotypic correlation due to rater specific views, or if the lower correlation based on ratings from different informants underestimate the true phenotypic correlation as a result of increased non-shared environmental influences.

In conclusion, we showed that a little under half of the variation in children's inattentive behavior is persistent over situations and is rater and setting independent. The heritability of this common phenotype is quite high. Todd et al. (2001) have argued that only through careful phenotype refinement will the identification of genetic and environmental influences on complex traits be realized. Because Attention Problems, which are persistent over situations, may indicate more serious behavior problems than Attention Problems that are present in only one context, we believe that gene finding strategies should focus on this common phenotype.

REFERENCE LIST

Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.

Bartels, M., Oord van den, E. J. C. G., Hudziak, J. J., Rietveld, M. J. H., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2004). Genetic and environmental mechanisms underlying stability and change in problem behaviors at ages 3, 7, 10, and 12. *Developmental Psychology, 40,* 852-867.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2004a). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research, 7,* 659-669.

Derks, E. M., Hudziak, J. J., Beijsterveldt van, C. E. M., Dolan, C. V., & Boomsma, D. I. (2004b). A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. *Behavior Genetics, 34,* 571-583.

Derks, E. M., Hudziak, J. J., and Boomsma, D. I. (in press). Genetics of ADHD, hyperactivity, and attention problems. *Handbook of Behavior Genetics*.

Eaves, L. J. (1976). A model for sibling effects in man. *Heredity, 36,* 205-214.

Eaves, L. J., Rutter, M., Silberg, J. L., Shillady, L., Maes, H. H., & Pickles, A. (2000). Genetic and environmental causes of covariation in interview assessments of disruptive behavior in child and adolescent twins. *Behavior Genetics, 30,* 321-334.

Eaves, L. J., Silberg, J. L., Meyer, J. M., Maes, H. H., Simonoff, E., Pickles, A. et al. (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 Psychology and Psychiatry and Allied Disciplines, 38,* 965-980.

Hewitt, J. K., Silberg, J. L., Neale, M. C., Eaves L.J., & Erickson, M. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics, 22,* 293-317.

Hudziak, J. J., Copeland, W., Stanger, C., & Wadsworth, M. (2004). Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 1299-1307.

Hudziak, J. J., Derks, E. M., Althoff, R. R., Rettew, D. C., & Boomsma, D. I. (2005). The genetic and environmental contributions to attention deficit hyperactivity disorder as measured by the Conners' Rating Scales - Revised. *American Journal of Psychiatry, 162,* 1614-1620.

Hudziak, J. J., Rudiger, L. P., Neale, M. C., Heath, A. C., & Todd, R. D. (2000). A twin study of inattentive, aggressive and anxious/depressed behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 469-476.

Kuntsi, J. & Stevenson, J. (2001). Psychological mechanisms in hyperactivity: II The role of genetic factors. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 211-219.

Leeuwen van, M., Berg van den, S. M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2005). Effects of twin separation in primary school. *Twin Research and Human Genetics, 8,* 384-391.

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

Martin, N., Scourfield, J., & McGuffin, P. (2002). Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *British Journal of Psychiatry, 180,* 260-265.

Nadder, T. S. & Silberg, J. L. (2001). Comparison of multiple measures of ADHD symptomatology: a multivariate genetic analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 475-486.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA: Department of Psychiatry.

Plomin, R. (1982). Childhood temperament. In B.Lahey & A. Kazdin (Eds.), *Advances in clinical child psychology*. New York: Plenum.

Pulkkinen, L., Kaprio, J., & Rose, R. J. (1999). Peers, teachers and parents as assessors of the behavioural and emotional problems of twins and their adjustment: the Multidimensional Peer Nomination Inventory. *Twin Research, 2,* 274-285.

Rietveld, M. J., Posthuma, D., Dolan, C. V., & Boomsma, D. L. (2003a). ADHD: sibling interaction or dominance: an evaluation of statistical power. *Behavior Genetics, 33,* 247-255.

Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2003b). Heritability of attention problems in children: I. Cross-sectional results from a study of twins, age 3-12 years. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 117B,* 102-113.

Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2004). Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 577-588.

Rietveld, M. J. H., Valk van der, J. C., Bongers, I. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D. I. (2000). Classification of twin zygosity. *Twin Research, 3,* 134-141.

Saudino, K. J., Ronald, A., & Plomin, R. (2005). The etiology of Behavior problems in 7-year-old twins: substantial genetic influence and negligible shared environmental influence for parent ratings and ratings by same and different teachers. *Journal of Abnormal Child Psychology, 33,* 113-130.

Saudino, K. J., Wertz, A. E., Gagne, J. R., & Chawla, S. (2004). Night and day: Are siblings as different in temperament as parents say they are? *Journal of Personality and Social Psychology, 87,* 698-706.

Sherman, D. K., McGue, M. K., & Iacono, W. G. (1997). Twin concordance for attention deficit hyperactivity disorder: A comparison of teachers' and mothers' reports. *American Journal of Psychiatry, 154,* 532-535.

Simonoff, E., Pickles, A., Hervas, A., Silberg, J. L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine, 28,* 825-837.

Thapar, A., Harrington, R., Ross, K., & McGuffin, P. (2000). Does the definition of ADHD affect heritability? *Journal of American Academy of Child and Adolescent Psychiatry, 39,* 1528-1536.

Todd, R. D., Rasmussen, E. R., Neuman, R. J., Reich, W., Hudziak, J. J., Bucholz, K. K. et al. (2001). Familiality and heritability of subtypes of attention deficit hyperactivity disorder in a population sample of adolescent female twins. *American Journal of Psychiatry, 158,* 1891-1898.

Towers, H., Spotts, E., Neiderhiser, J. M., Hetherington, E. M., Plomin, R., & Reiss, D. (2000). Genetic and environmental influences on teacher ratings of the Child Behavior Checklist. *International Journal of Behavioral Development, 24,* 373-381.

Ende van der, J. & Verhulst, F. C. (2005). Informant, gender and age differences in ratings of adolescent problem behavior. *European Child and Adolescent Psychiatry*.

Verhulst, F. C., Ende, J., & Koot, H. M. (1997). *Handleiding voor de Teacher's Report Form (TRF)*. Rotterdam: Afdeling Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/ Erasmus Universiteit Rotterdam.

Verhulst, F. C., Ende van der, J., & Koot, H. M. (1996). *Handleiding voor de CBCL/4-18*. Rotterdam: Afdeling Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.

Vierikko, E., Pulkkinen, L., Kaprio, J., & Rose, R. J. (2004). Genetic and environmental influences on the relationship between aggression and hyperactivity-impulsivity as reated by teachers and parents. *Twin Research, 7,* 261-274.

CHAPTER

11

GENETIC AND ENVIRONMENTAL INFLUENCES ON THE RELATION BETWEEN ATTENTION PROBLEMS AND ATTENTION DEFICIT HYPERACTIVITY DISORDER

This paper is submitted as Eske M. Derks, James J. Hudziak, Conor V. Dolan, Toos C.E.M. van Beijsterveldt, Frank C. Verhulst, Dorret I. Boomsma. Genetic and environmental influences on the relation between attention problems and attention deficit hyperactivity disorder. *Behavior Genetics, submitted.*

ABSTRACT

The assessment of symptoms of ADHD in children is usually established by the completion of a clinical interview with mothers or through the collection of checklists and questionnaires. The symptom scores obtained with different instruments are correlated although the association is less than perfect. The aim of the present study is to investigate the extent to which individual differences in the scores on different instruments are due to the same genetic and environmental influences. Maternal ratings were collected on 10916 twins from 5458 families. Child Behavior Checklist (CBCL) ratings were available for 10018, 6565, and 5780 twins at the ages 7, 10, and 12, respectively. The Conners Rating Scale (4887 twins) and the DSM interview (1006 twins) were completed at age 12. Genetic and environmental influences on the variance and covariance were obtained with Structural Equation Modeling. The phenotypic correlations of the three instruments range between .45 and .77. The variances and covariances of the five instruments were mainly explained by genetic influences. The genetic correlations of the data collected at age 12 varied between .61 and 1.00. The genetic overlap between behavior checklist scores and the DSM-IV diagnosis of ADHD is high. There are important clinical and research implications to these findings which are presented.

INTRODUCTION

As is the case for all psychiatric disorders, the diagnosis of attention deficit hyperactivity disorder (ADHD) is not based on a specific pathological agent, such as a microbe, a toxin, or a genetic mutation, but rather on the collection of signs and symptoms and evidence of impairment, that occur together more frequently than expected by chance (Todd et al., 2005). The presence of these symptoms is usually established by direct observation, or by the completion of a clinical interview or questionnaire by the parent or teacher of the child. Instruments show variation in the symptoms, how the symptoms are collected (checklist or interview), and how they are rated (yes/no or on a likert scale). In the present paper, we investigated the extent to which individual differences in the scores on different instruments are due to the same genetic and environmental influences. The focus is on three widely used instruments: the Child Behavior Checklist (CBCL; Achenbach, 1991), the Conners Parent Rating Scale-Revised:Short version (CPRS-R:S; Conners, 2001), and the Diagnostic and Statistical Manual of Mental Disorders-4th edition (DSM-IV; American Psychiatric Association, 1994).

The CBCL-Attention Problem scale (CBCL-AP) was developed by means of factor analyses, and includes eleven items. The psychometric properties and methods to establish the reliability of the syndrome is discussed in detail elsewhere (Achenbach, 1991). Despite its name, the scale does not assess problems related exclusively to AP. The CBCL has sex and age specific norms which are useful in assessing whether a child is at risk of ADHD. The CPRS-R:S ADHD-index (ADHD-I) comprises the twelve best items for distinguishing children with ADHD from children without ADHD as assessed by the DSM (Conners, 2001). As for the CBCL, norm scores are provided that allow the clinician to compare a raw score with gender and age specific T scores to determine whether or not an individual child is at risk for ADHD. DSM-IV ADHD is assessed on the basis of 18 symptoms: nine relate to inattention, and nine relate to hyperactivity/impulsivity. In the DSM framework, ADHD is viewed as a categorical trait: i.e., children either do or do not meet criteria for ADHD. The norms for clinical diagnosis do not vary as a function of sex or age of the child.

Although the three approaches focus on different symptoms and are based on distinct assumptions, the outcome measures of the two instruments are stongly related. CBCL-AP scores predict the presence of ADHD (Gould et al., 1993; Chen et al., 1994; Eiraldi et al., 2000; Lengua et al., 2001; Sprafkin et al., 2002; Hudziak et al., 2003). In a non-referred sample enriched for ADHD, about 50% percent of the children with a high CBCL-AP score were diagnosed with ADHD compared to 3% percent of the children with a low CBCL-AP score (Derks et al., 2006). Although these numbers imply a good convergence between the two measures, clearly the relation is less than perfect. The Conners ADHD-I index was developed for assessing children at risk for ADHD based on a DSM-IV diagnosis (Conners, 2001). Conners (2001) showed that the Conners ADHD-I is a good screening instrument for DSM-IV ADHD with a sensitivity of 100%, a specificity of 92.5%, and an overall correct classification rate of 96.3%. As far as we know, the relation between CBCL-AP and the Conners ADHD-I has not been studied, but given that they are both related to DSM-IV ADHD, it is likely that these two measures are also correlated.

Genetic studies of psychiatric disorders are complicated by the lack of clear diagnostic tests (Hudziak, 2001). Heritability estimates in epidemiological genetic studies, and the results of gene-finding studies may vary as a consequence of the instrument that is used to assess ADHD. Although a number of papers have established the convergence between CBCL-AP and ADHD, the causational factors underpinning this relationship remain unclear. Is it the result of genetic overlap, environmental overlap, or both? This is an important question, which may affect the progress in gene finding studies. If variation in alternative measures of ADHD is explained by different genes, we would expect disagreement in

the results of studies using different instruments. If the same genes explain variation in these measures, the data from studies using different instruments may be combined in order to increase statistical power (Boomsma, 1996; Boomsma & Dolan, 1998).

Genetic and environmental influences on individual differences in behavior can be identified with genetically informative data. The classical twin design, involving the comparison of monozygotic (MZ) and dizygotic (DZ) twins, is used most frequently to this end. Previous studies of the concordance rates in MZ and DZ twins showed that genetic influences explain between 55 to 89% of the variation in clinical diagnoses of ADHD (Eaves et al., 1997; Sherman et al., 1997). Shared environmental influences were found to be absent. Likewise, variation in CBCL-AP scores is mainly explained by genetic influences, and not by shared environmental influences (Rietveld et al., 2003; Hudziak et al., 2000; Gjone et al., 1996). Interestingly, in parent ratings, but not in teacher ratings, the DZ twin concordances and correlations are lower than would be expected under a purely additive genetic model. For example, in maternal structured interview reports, the concordance rate is .67 in MZ twins compared to .00 in DZ twins (Sherman et al., 1997). Similarly, in CBCL ratings, the DZ twin correlations are less than half the MZ correlations (Rietveld et al., 2003). In the literature, two explanations are offered for these low correlations. Firstly, the DZ correlation can be less than half the MZ correlation due to the presence of non-additive genetic effects (i.e., genetic dominance) (Lynch & Walsh, 1998). Secondly, the low DZ correlation may be explained by social interaction effects, which may be the result of interaction among siblings (i.e., the behavior of a twin influences the behavior of the other twin) or rater bias (i.e., the behavior of a twin is compared to the behavior of the other twin) (Eaves, 1976; Carey, 1986; Boomsma, 2005). In previous studies, support was found both for the presence of genetic dominance (Rietveld et al., 2003; Martin et al., 2002) and sibling interaction (Simonoff et al., 1998; Kuntsi & Stevenson, 2001; Vierikko et al., 2004; Eaves et al., 1997).

Although a number of studies have focused on the genetic and environmental influences on either AP or ADHD, as far as we know, only one study included multivariate analyses. Nadder & Silberg (2001) collected data in a sample of 735 male and 819 female same-sex twin pairs, aged 8 to 16 years. They modelled the genetic influences on nine measures of ADHD symptomatology, including maternal and paternal DSM-III-R interview data (three dimensions: hyperactivity, inattention and impulsivity), the Rutter Parental Scale, the CBCL, and a questionnaire completed by the twin's teacher. The aim of this study was to determine whether overactivity, inattention, and impulsivity reflect the same underlying genetic liability while taking method variance into account. In males, 23.7-70.1% of the genetic variation was explained by a common factor that loads on all nine indicators. A second and third factor loaded on the three dimensions of the maternal and paternal interview data, respectively. The remaining variance (0.0-65.7%) was explained by factors that were specific to each measure. In females, there was also one factor common to all indicators (explaining 16.2-60.2% of the variation), and a second and third factor which loaded on the three dimensions of the interview data. In contrast to the males, a fourth factor loaded on the three behavioral questionnaires. This factor explained 12.3-46.2% of the genetic variation. In total, measurement specific factors explained 0.0-73.0% of the genetic variance. These analyses showed that variation in the nine indicators of ADHD is not completely explained by a single common genetic factor. However, it is not clear if the specific genetic factors reflect instrument variance (e.g., diagnostic interview versus questionnaires), or rater variance (e.g., maternal versus paternal ratings).

The purpose of the present paper is to investigate the genetic and environmental influences on individual differences in CBCL-AP, CPRS-R:S ADHD-I, and DSM-IV ADHD. Three questions are addressed. First, what are the phenotypic correlations between the three instruments? Second, to what extent are the phenotypic correlations explained by genes and environment? Third, to what extent do the genetic and environmental influences,

which explain individual differences in the three instruments, overlap? Although the last two questions may appear to be similar, they do differ. For example, variation in CBCL, Conners and DSM scores may be explained by the same set of genes, but these genes may explain only a small proportion of the covariance between the measures (i.e., genes overlap completely but explain only a small portion of the phenotypic covariance). Alternatively, variation in CBCL, Conners, and DSM scores may be explained by genes that only partially overlap but the genes that do overlap may explain a large proportion of the covariance (i.e., genes overlap only partially, but do explain a large proportion of the phenotypic covariance).

METHODS

Subjects

This study is part of an ongoing longitudinal twin study in the Netherlands. The subjects were all registered at birth with the Netherlands Twin Register (Boomsma et al., 2002). Mothers of the registered twin-pairs receive the CBCL and the CPRS at the ages 7, 10, and 12 years. A subsample of the twins was selected based on their longitudinal CBCL scores. The mothers of these pairs completed a diagnostic interview. The twins, with an age range of 10 to 13 years (mean age=11.71; sd=.77) at the time of the interview were born between 1989 and 1994. The mean time-span between the completion of the interview and the questionnaires was 4.42 (sd=.75), 1.82 (sd=.73), and -.84 (sd=.63) for the questionnaires filled in at age 7, age 10, and age 12, respectively.

At least one measurement is available for 10916 twins from 5458 families. CBCL ratings were available in 10018 twins at age 7, 6565 twins at age 10, and 5780 twins at age 12. CPRS-R:S ratings were available for 4887 twins at age 12, and DSM-IV interviews were available for 1006 twins. The number of available questionnaires decreases over time as a result of the longitudinal character of the study (i.e., a number of children in the study had yet to reach the age of 12).

Zygosity diagnosis was based on DNA in 674 same-sex twin pairs. In the remaining same-sex pairs, zygosity was assessed using a 10-item questionnaire. Zygosity determination using this questionnaire is almost 95% accurate (Rietveld et al., 2000). Of the 5458 twin-pairs, there were 898 monozygotic male (MZM) pairs, 888 dizygotic male (DZM) pairs, 1005 monozygotic female (MZF) pairs, 844 dizygotic female (DZF) pairs, and 1823 dizygotic opposite sex (DOS) pairs.

Selection for the diagnostic interview

For the diagnostic interview, subjects were selected on the basis of their standardized maternal CBCL ratings (T-scores; Mean=50, SD=10) at the ages 7, 10, and 12 years. Subjects were excluded if maternal ratings were available only at one time-point, or if they suffered from a severe handicap, which disrupted daily functioning. Twin-pairs were selected if at least one of the twins scored high on AP (affected pairs), or if both twins scored low on AP (control pairs). A high score was defined as a T-score above 60 at all available time-points (age 7, 10, and 12 years) and a T-score above 65 at least once. A low score was defined as a T-score below 55 at all available time-points. The control pairs were matched with the affected pairs on the basis of sex, cohort, maternal age, and Social Economic Status (SES). T-scores were computed in boys and girls separately. In other words, girls were selected if they scored low or high compared to other girls, and boys were selected if they scored low or high compared to other boys. This procedure resulted in the selection of similar numbers of boys (N=499) and girls (N=507).

Measures

The Child Behavior Checklist (CBCL) (Achenbach, 1991) is a standardized questionnaire designed for parents to report the frequency and intensity of their children's behavioral and emotional problems as exhibited in the past six months. It consists of 120 items that measure problem behavior. The items are rated on a 3-point scale ranging from "not true", "somewhat or sometimes true", to "very true or often true". The Attention Problem scale contains 11 items. The two week test-retest correlation and the internal consistency of this scale are .83 and .67, respectively (Verhulst et al., 1996). In the statistical analyses, we included the CBCL ratings at the ages 7, 10, and 12 years in order to correct for the selection, as explained below.

The Conners' Parent Rating Scale-Revised is a widely used instrument to assess behavior problems (CPRS-R; Conners, 2001; Conners et al., 1998). The short version contains 28 items. The items are rated on a 4-point scale ranging from "not true at all" to "very much true". The ADHD-index, which was used in the present study, comprises the best twelve items for distinguishing children with ADHD from children without ADHD as assessed by the DSM-IV (American Psychiatric Association, 1994; Conners, 2001). The internal consistency of this scale at age 12-14 years is .94 in boys and .91 in girls. The 6-8 weeks test-retest correlation is .72.

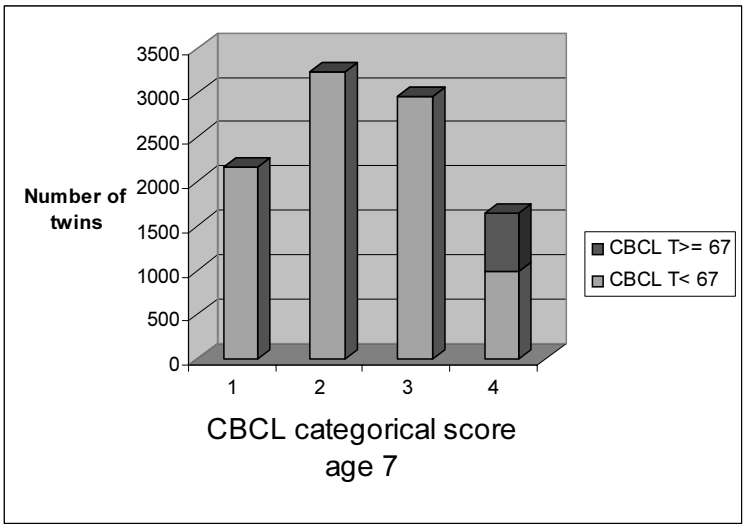
The Diagnostic Interview Schedule for Children (DISC) (Shaffer et al., 1993) is a structured diagnostic interview. It can be used to assess the presence of DSM-IV diagnoses, including ADHD. The Dutch translation is due to Ferdinand and Ende van der (1998). The mothers of twins were interviewed by 10 experienced research assistants to determine which symptoms of ADHD were displayed by the twins during the last year. We analyzed the total number of symptoms. The sumscore with a range of 0 to 18 was transformed into an ordinal variable with four categories to correct for the skewness and kurtosis of the distribution of the sum scores as explained below. The four categories were: i) not affected (0 symptoms); ii) mildly affected (1-2 symptoms); iii) moderately affected (3-5 symptoms); and iv) highly affected (more than 6 symptoms). The use of this four category variable provides greater resolution, and so better statistical power than the use of a dichotomous variable (ADHD absent vs ADHD present).

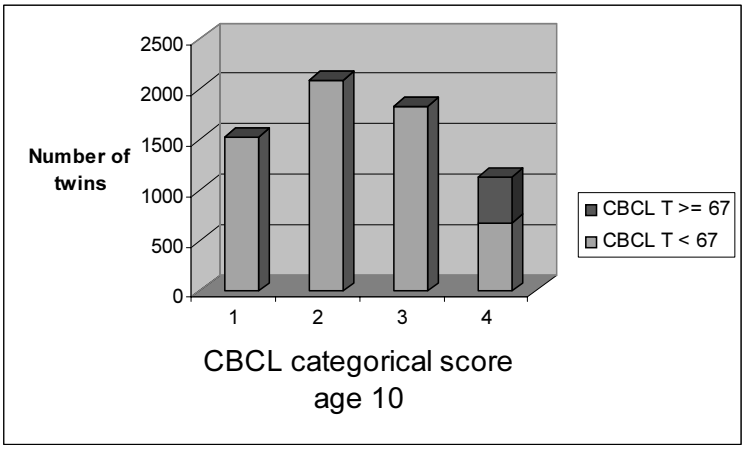
Statistical analyses

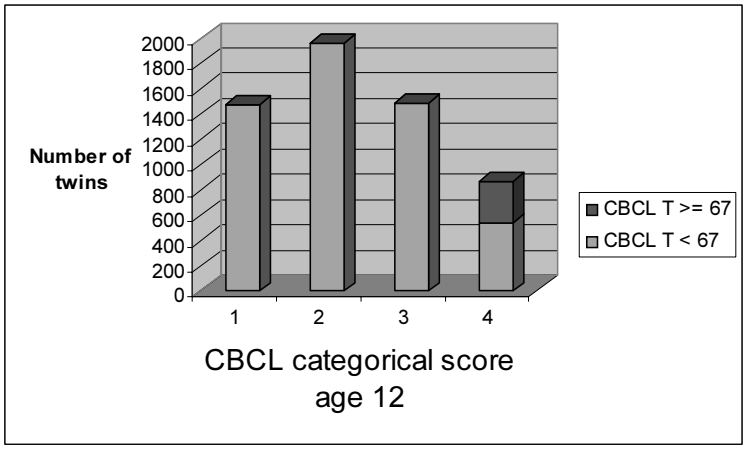
Transformation to categorical data

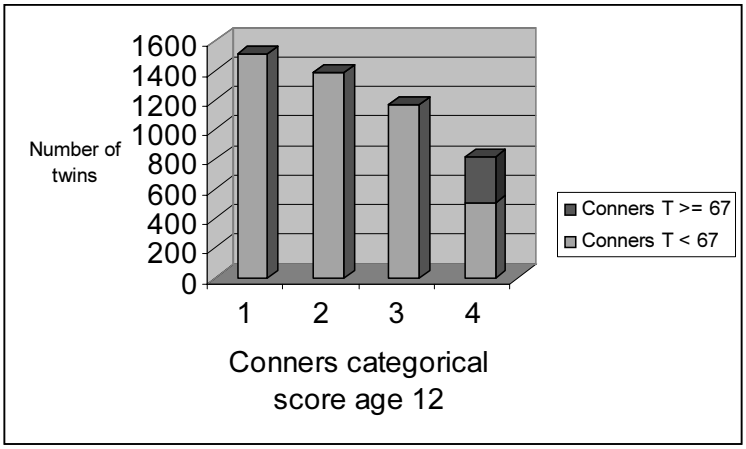
The distributions of the CBCL, CPRS-R:S, and DSM symptom data are characterized by excessive skewness and kurtosis. Derks, Dolan, & Boomsma (2004) showed that bias in parameter estimates due to non-normality of the data may be avoided by using categorical data analysis. In this approach, a liability threshold model is applied to the ordinal scores (Lynch & Walsh, 1998). It is assumed that a person is "unaffected", if his or her liability is below a certain threshold, and that he or she is "affected", if his or her liability is above this threshold. In the present paper, the scores were recoded in such a way that three thresholds divide the latent liability distribution into four categories. The prevalences are more or less similar in each of the four categories. Figures 1a-e show the relation between the categorical scores and the respective clinical cutpoints of the CBCL (Verhulst et al., 1996), the Conners ADHD-I (Conners, 2001), and the DSM-IV diagnosis of ADHD (American Psychiatric Association, 1994).

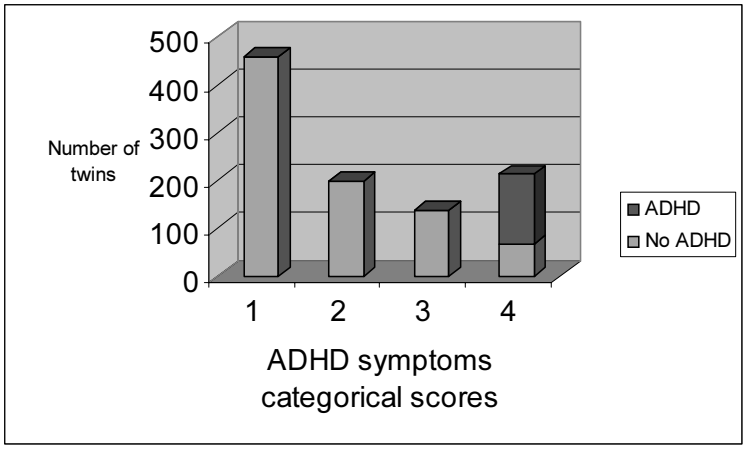
Figure 1. a-e: Overlap between the categorical sumscores which are used in the present study and the clinical cutpoints











Correcting for the selection

Diagnostic interview data were only collected for a subsample of the twins. The probability of selection for the interview depends on a measured variable, namely the twin's CBCL scores. In the terminology of Little & Rubin (2002), the data are Missing At Random (MAR). Given that the data are MAR, unbiased parameter estimates can be obtained by full information (i.e., raw data) maximum likelihood estimation of the parameters in a statistical model that includes the variables that were used for selection. We therefore included the CBCL ratings obtained at the ages 7, 10, and 12 years in the statistical analyses. All twins for whom at least one measure is available are included in the analysis Derks et al. (in press).

Prevalences

In order to investigate if the prevalences of AP and ADHD depend on the twin's sex or zygosity, we performed χ^2 tests with the five ordinal measures as dependent variables and sex and zygosity as independent variables.

Genetic Modeling

Genetic and environmental influences on the ADHD scores were estimated using structural equation modeling. All model fitting was performed on raw data with MX (Neale et al., 2003), a statistical software package designed for conducting genetic analyses.

The influence of the relative contributions of genetic and environmental factors to individual differences in ADHD can be inferred from the different levels of genetic relatedness of MZ and DZ twins (Plomin et al., 2001). The variance may be due to additive genetic

effects (A), dominant genetic effects (D) or shared environmental effects (C), and non-shared environmental (E) effects. The genetic effects (A and D) correlate 1 in MZ twins as they are genetically identical. In DZ twins, A correlates .5, because DZ twins share on average half of their segregating genes. The effects due to dominance correlate .25 in DZ twins. C correlates 1 in both MZ and DZ twins. E or non-shared environmental effects are, by definition, uncorrelated. All uncorrelated measurement error, if present, is also absorbed in the E term. Note that estimating C and D at the same time is not possible in a design using only data from MZ and DZ twins reared together. If the correlations of DZ twins are less than half the correlations of MZ twins, which is the case for maternal ratings of attention problems and ADHD, D is included in the genetic model. The proportion of the variation accounted for by heritability or environmental influences is calculated by calculating the ratio of variance due to A, D, or E to the total phenotypic variance. For instance let a, d, and e denote the regression coefficients in the regression of the phenotype on A, D, and E, respectively. The variance due to A is then a2, and the (narrow-sense) heritability is calculated as $a^2/(a^2 + d^2 + e^2)$.

Social interactions may be an additional source of variance. Social interaction effects lead to differences in variances in MZ and DZ twins in continuous data (Carey, 1986). Using ordinal data, the presence of an interaction component can be tested by equating the prevalences of AP/ADHD between MZ and DZ twins. The absence of significant prevalence differences suggests that the presence of sibling interaction or rater bias is implausible.

The full ADE Cholesky decomposition that was fitted to the multivariate data is shown for an individual twin in Figure 2. Five factors explain the variation and covariation of the CBCL scores at the ages 7, 10, and 12 years, and on the CPRS-R:S and DSM symptom scores. The first factor loads on all five traits, the second factor loads on traits 2-5, the third factor loads on traits 3-5, the fourth factor loads on traits 4-5, and the fifth factor loads on trait 5 only. In the Cholesky decomposition, the effects of A, D, and E are represented by a triangular matrix of regression coefficients, or factor loadings. For example, with the factors in columns and the variables in rows, the factorloadings of A can be denoted as a matrix:

$$\begin{pmatrix} a_{11} & & & & \\ a_{21} & a_{22} & & & \\ a_{31} & a_{32} & a_{33} & & \\ a_{41} & a_{42} & a_{43} & a_{44} & \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}$$

This matrix multiplied by its transpose results in the additive genetic variance-covariance matrix:

$$\begin{pmatrix} a_{11}^2 & & & & \\ a_{21}a_{11} & a_{22}^2 + a_{21}^2 & & & \\ a_{31}a_{11} & a_{32}a_{22} + a_{31}a_{21} & a_{33}2 + a_{32}^2 + a_{31}^2 & & \\ a_{41}a_{11} & a_{42}a_{22} + a_{41}a_{21} & a_{43}a_{33} + a_{42}a_{32} + a_{41}a_{31} & a_{44}^2 + a_{43}^2 + a_{42}^2 + a_{41}^2 & \\ a_{51}a_{11} & a_{52}a_{22} + a_{51}a_{21} & a_{53}a_{33} + a_{52}a_{32} + a_{51}a_{31} & a_{54}a_{44} + a_{53}a_{43} + a_{52}a_{42} + a_{51}a_{41} & a_{55}^2 + a_{54}^2 + a_{53}^2 + a_{52}^2 + a_{51}^2 \end{pmatrix}$$

Dividing this matrix by the implied phenotypic variance-covariance matrix provides the proportion of variances and covariances explained by additive genetic effects; standardizing it provides the genetic correlation matrix where the correlations indicate the overlap of genetic effects across time and instrument.

Because the number of twins, in whom interview data are available, is relatively small, and sex differences in heritability are usually not found, the data from male and female twins were combined. To allow for prevalence differences between boys and girls, sex was included as a covariate on the thresholds. The type-I error rate of all statistical tests was set at .01 (rather than .05) to accommodate multiple testing.

RESULTS

Descriptives

The prevalences were compared between boys and girls and between MZ and DZ twins. Boys showed significantly more problems than girls ($\chi^2(2)=458$, $p<.001$). Zygosity did not effect the prevalences of the CBCL, Conners, and DSM scores ($\chi^2(30)=48$, $p=.018$). Because of the absence of prevalence differences in MZ and DZ twins, no social interaction effects were included in the genetic model.

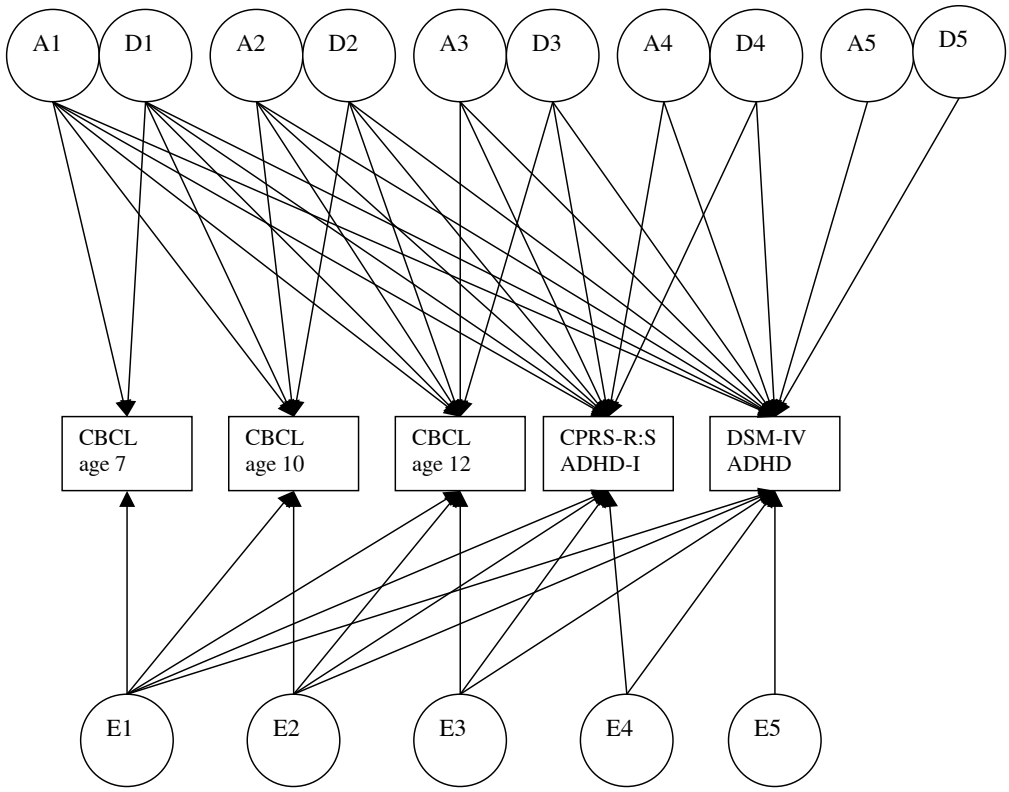
Twin correlations

The polychoric correlations between the five measurements are shown in Table 1 for MZ and DZ twins. The MZ (DZ) twin correlations are reported above (below) the diagonal. As expected, the phenotypic correlations (i.e., the correlation between traits within the same individual) are similar in first and second born twins and in MZ and DZ twins. The correlations are in the range of .45 to .77, and do not seem to differ much when comparing different assessment methods (e.g., CBCL questionnaire vs. clinical interview) and similar assessment methods (e.g., CBCL questionnaire vs. CPRS-R:S questionnaire). The fact that the cross-twin and the cross-trait cross-twin correlations are higher in MZ than DZ twins indicates that genetic influences contribute to the variance of the three measures and the covariance between them.

Genetic analyses

A Cholesky decomposition that included additive genetic influences (A), dominant genetic influences (D), and non-shared environmental influences (E) was fit to the data. The full ADE cholesky model provides a good fit to the data ($\chi^2(50)=59$, $p=.180$). The five additive genetic, and non-shared environmental factors in the Cholesky decomposition all showed substantial loadings on the five measurements. In contrast, the loadings of the dominant genetic factors suggested that a single factor may fit the dominant genetic covariance structure well. The fit of the model including a Cholesky decomposition for A and E, and a single factor for D provided a good fit in comparison with the full Cholesky ($\chi^2(10)=2.07$, $p=1.00$). The covariance structure of D did not include specific variances. This means that the covariance matrix has rank one, and that the correlations (obtained by standardizing the covariance matrix of D) were all one. Table 2 shows the standardized influences of A, D, and E on the variance and covariance of the five measurements. On the three diagonals of the five by five tables of A, D, and E, one can find the standardized influences on the variances. For example, the variance of the CBCL rating at age 7, is for 47% explained by A, for 30% by D, and for 23% by E. The total, or broadsense, heritability of the CBCL at age 7, is 47 + 30 = 77%. Likewise, the heritabilities of the remaining four measures are 78%, 76%, 85%, and 64% for CBCL age 10, CBCL age 12, CPRS-R:S ADHD-I, and DSM-IV ADHD, respectively. Thus, the heritability is high, irrespective of measurement instrument or age.

Figure 2. Cholesky decomposition of the genetic and environmental influences on the variances and covariances of five ratings of ADHD and attention problems



Note: Circles represent latent, unmeasured variables; Squares represent observed variables; A=additive genetic effects; D=dominant genetic effects, E=non-shared environmental effects; CBCL=Child Behavior Checklist; CPRS-R:S ADHD-I=Conners Parent Rating Scale-Revised: Short version ADHD-index; DSM=Diagnostic Statistical Manual of Mental Disorders

Table 1. Polychoric correlations in monozygotic (above diagonal) and dizygotic (below diagonal) twins

		First-born					Second born				
		CBCL 7	CBCL 10	CBCL 12	CPRS	DSM	CBCL 7	CBCL 10	CBCL 12	CPRS	DSM
	CBCL age 7	1	.66	.62	.51	.59	.76	.54	.49	.45	.45
	CBCL age 10	.70	1	.69	.61	.59	.56	.77	.58	.53	.48
First-born	CBCL age 12	.63	.74	1	.71	.57	.48	.54	.75	.58	.53
	CPRS-R:S	.56	.68	.77	1	.60	.46	.55	.62	.84	.51
	DSM	.51	.55	.59	.68	1	.34	.41	.46	.46	.64
Second born	CBCL age 7	.31	.22	.18	.15	.04	1	.66	.63	.52	.46
	CBCL age 10	.22	.35	.22	.21	.01	.66	1	.71	.64	.59
	CBCL age 12	.21	.28	.34	.24	.13	.60	.72	1	.75	.58
	CPRS-R:S	.22	.27	.28	.38	.08	.49	.64	.74	1	.60
	DSM	.11	.16	.11	.07	.13	.45	.63	.67	.58	1

Note:CBCL=Child Behavior Checklist; CPRS=Conners Parent Rating Scale-Revised: Short version ADHD-index; DSM=Diagnostic Statistical Manual of Mental Disorders

Table 2. Standardized genetic and environmental influences on the variances and covariances of five ratings of ADHD and attention problems

	A					D					E				
	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM
CBCL age 7	.47					.30					.23				
CBCL age 10	.44	.55				.39	.23				.17	.22			
CBCL age 12	.40	.52	.57			.39	.29	.19			.21	.19	.24		
CPRS-R:S	.39	.52	.55	.62		.49	.35	.27	.22		.12	.13	.18	.16	
DSM	-.07	.14	.29	.23	.10	.84	.61	.52	.55	.55	.23	.25	.19	.22	.35

Note:CBCL=Child Behavior Checklist; CPRS-R:S=Conners Parent Rating Scale-Revised: Short version ADHD-index; DSM=Diagnostic Statistical Manual of Mental Disorders; A=additive genetic effects; D=dominant genetic effects; E=non-shared environmental effects

On the off-diagonals of Table 2, one can find the influences of A, D, and E on the covariance between the measurements. The most interesting comparison is between the data that were collected at approximately the same time. The covariance between the CBCL at age 12 and the DSM is for 81% explained by genetic effects (29% A and 52% D) and for 19% by E. The covariance between the Conners and the DSM is for 77% explained by genetic effects (23% A and 55% D), while the covariance between the CBCL age 12 and Conners is for 82% explained by genetic effects (55% A and 27% D).

Table 3 includes the genetic and environmental correlation matrices, which represent the overlap between the genetic and environmental influences on the five measurement instruments. Except the correlation of -.16 for the CBCL at age 7 and the DSM interview, the additive genetic correlations range between .36 and .78. The low correlation of -.16 may seem surprising, but is not very meaningful, because the covariance between the CBCL at age 7 and the DSM interview is not explained by A, but rather by D. All dominant genetic correlations are 1, which is a direct result from the fact that a one-factor model without specifics explained the dominant genetic variance. The non-shared environmental correlations are in the range of .33 to .67. Although the non-shared environmental effects play a relatively small role in explaining the variance and covariance of the five measures, the influences that do play a role show large overlap.

Table 3. Genetic and environmental correlations of five ratings of ADHD and attention problems

	A					D					E				
	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM
CBCL age 7	1.0					1.0					1.0				
CBCL age 10	.59	1.0				1.0	1.0				.50	1.0			
CBCL age 12	.48	.67	1.0			1.0	1.0	1.0			.54	.59	1.0		

Note:CBCL=Child Behavior Checklist; CPRS-R:S=Conners Parent Rating Scale-Revised: Short version ADHD-index; DSM=Diagnostic Statistical Manual of Mental Disorders; A=additive genetic effects; D=dominant genetic effects; E=non-shared environmental effects

Table 3. Genetic and environmental correlations of five ratings of ADHD and attention problems

	A					D					E				
	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM	CBCL 7	CBCL 10	CBCL 12	CPRS-R:S	DSM
CPRS-R:S	.38	.57	.70	1.0		1.0	1.0	1.0	1.0		.33	.47	.67	1.0	
DSM	-.16	.36	.78	.61	1.0	1.0	1.0	1.0	1.0	1.0	.38	.53	.40	.57	1.0

Note:CBCL=Child Behavior Checklist; CPRS-R:S=Conners Parent Rating Scale-Revised: Short version ADHD-index; DSM=Diagnostic Statistical Manual of Mental Disorders; A=additive genetic effects; D=dominant genetic effects; E=non-shared environmental effects

DISCUSSION

The aim of this study was to determine the aetiology of the covariance between three different instruments, which are commonly used to assess ADHD, attention problems, and hyperactivity. The instruments under consideration are two scales based on items from questionnaires (CBCL-AP, and Conners ADHD-I), and a DSM-IV ADHD interview. First, we considered the phenotypic correlations. Second, we estimated the proportion of covariance that is explained by genetic influences. Third, we investigated the extent to which the sets of genes that explain variation in these three measures overlap. This is the first study that includes multivariate genetic analyses of behavior rating scales and DSM-IV interview data collected in a large sample of twins of approximately the same age. The CBCL scores collected at age 7 and 10 years were only included to correct for the selection. In the discussion, we focus mainly on the CBCL, Conners and DSM interview data, which were collected at age 12 years.

The phenotypic correlation between CBCL-AP and the Conners ADHD-I was high (r=.75). The correlations between the CBCL and the DSM and between the Conners and the DSM were slightly lower (r=.62). These lower correlations can both be the result of the different time-points at which the behavior checklists and the DSM interview data were collected (the mean time-span between measurement occasions was .84 year) and the result of instrument or method variance (e.g., interview vs behavior checklists). The first explanation is supported by the fact that the correlation between the Conners at age 12 and the CBCL at age 10 is lower than the correlation between the Conners and CBCL, which were completed at the same age. However, the fact that the correlation between the CBCL and Conners with a two year time-span (r=.65) is similar to the correlation between the behavior checklists and the DSM data with slightly less than one year time span (r=.62) indicates that instrument variance may also be a factor. As noted above, the AP scale of the CBCL questions relate to both inattention and hyperactivity/impulsivity. The fact that the correlation between the Conners ADHD-I and DSM-IV ADHD is identical to the correlation between CBCL-AP and DSM-IV ADHD implies that the Conners and the CBCL measure ADHD equally well. The description of the eleven item CBCL scale as an AP scale seems to be too limited, because both the item content and the current results suggest that the CBCL also signals problems related to hyperactivity/impulsivity.

Although the phenotypic correlations provide an interesting insight in the similarities and dissimilarities of the quantitative and qualitative approaches towards child psychopathology, an important question concerns the aetiological influences on the covariance between the instruments. The covariances between the three instruments were mainly explained by genetic factors. Nonshared environmental influences played a relatively small role. This shows that the agreement between different approaches towards psychopathology is the result of genetic rather than environmental influences.

The third question concerned the overlap between the sets of genes that explained variation in the three instruments. This is particularly interesting, because high genetic correlations would imply that the detection of the specific genes that play a role for ADHD, does not depend much on the instrument that is used. At age 12, the additive genetic correlations of the CBCL, Conners, and DSM varied between .61 and .78, while the dominant genetic correlations could be constrained at 1. The non-shared environmental correlations are also quite high, and vary between .40 and .67. The dominant genetic correlations of 1 suggest that there is a subset of genes of which the effect is not instrument or age dependent. In contrast, the correlations of the additive genetic effects are high but less than perfect. This suggests that the influence of most genes with an additive effect are not sensitive to the particular instrument that is used, although there are some genes that explain variation only in a particular measurement (e.g., CBCL), but not in another (e.g., DSM).

The correlations of the additive and dominant genetic factors are not higher between behavior checklist ratings than between behavior checklist ratings and interview data. Instead, the higher phenotypic correlation of the Conners and CBCL is explained by higher non-shared environmental correlations. Because non-shared environmental influences also include measurement error, it seems likely that the greater covariance of the CBCL/Conners compared to the CBCL/DSM and Conners/DSM is the result of instrument variance that is shared between the CBCL and Conners, but not by the DSM. The discrepancy between behavior checklist and interview scores does not seem to be the result of a fundamentally different way to conceptualize problem behavior, but might be explained by the fact that measurement error is correlated when similar instruments are used to assess problem behavior.

What are the implications of the present findings for gene finding studies? Thus far, five groups have conducted genome-wide linkage scans in an attempt to find genomic regions which are involved in ADHD, but the results of these studies have been inconsistent. Linkage peaks with a LOD score above 2 ($p<\sim.002$) were reported at chromosomes 16p13 and 17p11 (Ogdie et al., 2003), chromosomes 7p and 15q (Bakker et al., 2003), chromosomes 4q, 8q, and 11q (Arcos-Burgos et al., 2004), chromosomes 5p and chromosome 12q (Hebebrand et al., 2006), and chromosomes 14q32 and 20q11 (Gayan et al., 2005). All these studies based diagnosis on DSM-IV (Ogdie et al., 2003; Bakker et al., 2003; Arcos-Burgos et al., 2004; Hebebrand et al., 2006) or DSM-III (Gayan et al., 2005) criteria. The discrepancy in the results of these five studies could be the result of a lack of statistical power. The present study showed that the genetic overlap between behavior checklist scores and the DSM-IV diagnosis of ADHD is high. This implies that the detection of genes which play a role for ADHD can be based on survey data. This will reduce the costs associated with the collection of the phenotypic data in comparison to the collection of diagnostic interviews. The resources which do not have to be used for the phenotyping can be used for the collection of genotypic data. An increased number of subjects can be genotyped and the statistical power to detect a QTL will be increased.

Limitations

The results of this study should be interpreted bearing in mind the following limitations. First, further study is required to investigate if the results of the current study, which was based on a Dutch population sample, generalize to population samples in the United States. Second, clinical diagnoses were based on structured diagnostic interviews with the mother. The results may be different when the assessment of ADHD is based on expert clinical diagnoses.

Clinical implications

Two general approaches towards the measurement of ADHD can be distinguished. In the DSM-IV framework, ADHD is viewed as a categorical trait. Using behavior checklists, children can show variation in a continuum from not affected at all to severely affected. The

current study shows that variation in DSM-IV symptoms, the CBCL-AP scale, and the Conners ADHD-I is explained mostly by genetic effects. The correlations between the genetic influences on variation in these three measurements of ADHD are very high. This implies that different measurements tap the same genetic liability.

REFERENCE LIST

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

American Psychiatric Association (1994). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorder: DSM-IV (4th ed.)*. Washington, DC: American Psychiatric Association.

Arcos-Burgos, M., Castellanos, F. X., Pineda, D., Lopera, F., Palacio, J. D., Palacio, L. G. et al. (2004). Attention-deficit/hyperactivity disorder in a population isolate: Linkage to loci at 4q13.2, 5q33.3, 11q22, and 17p11. *American Journal of Human Genetics, 75,* 998-1014.

Bakker, S. C., Meulen van der, E. M., Buitelaar, J. K., Sandkuijl, L. A., Pauls, D. L., Monsuur, A. J. et al. (2003). A whole-genome scan in 164 Dutch sib-pairs with Attention-Deficit Hyperactivity Disorder: suggestive evidence for linkage on chromosomes 7p and 15q. *American Journal of Human Genetics, 72,* 1251-1260.

Boomsma, D. I. (1996). Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. *Behavior Genetics, 26,* 161-166.

Boomsma, D. I. (2005). Sibling Interaction Effects. In B.Everitt & D. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (pp. 1831-1832). Chisester: John Wiley & Sons.

Boomsma, D. I. & Dolan, C. V. (1998). A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor scores. *Behavior Genetics, 28,* 329-340.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics, 16,* 319-340.

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the child behavior checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62,* 1017-1025.

Conners, C. K. (2001). *Conners' Rating Scales-Revised*. New York and Toronto: Multi-Health Systems, Inc.

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26,* 257-268.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research, 7,* 659-669.

Derks, E. M., Dolan, C. V., and Boomsma, D. I. (in press). Statistical power to detect genetic and environmental influences in the presence of data missing at random. *Twin Research and Human Genetics.*

Derks, E. M., Hudziak, J. J., Dolan, C. V., Ferdinand, R. F., & Boomsma, D. I. (2006). The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. *Comprehensive Psychiatry, 47,* 116-122.

Eaves, L. J. (1976). A model for sibling effects in man. *Heredity, 36,* 205-214.

Eaves, L. J., Silberg, J. L., Meyer, J. M., Maes, H. H., Simonoff, E., Pickles, A. et al. (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 Psychology and Psychiatry and Allied Disciplines, 38,* 965-980.

Eiraldi, R. B., Power, T. J., Karustis, J. L., & Goldstein, S. G. (2000). Assessing ADHD and comorbid disorders in children: the Child Behavior Checklist and the Devereux Scales of Mental Disorders. *Journal of Clinical Child Psychology, 29,* 3-16.

Ferdinand, R. F. & Ende van der, J. (1998). *Diagnostic Interview Schedule for Children IV Parent-version*. Rotterdam: Erasmus University Rotterdam, Department of Child and Adolescent Psychiatry.

Gayan, J., Willcutt, E. G., Fisher, S. E., Francks, C., Cardon, L. R., Olson, R. K. et al. (2005). Bivariate linkage scan for reading disability and attention-deficit/hyperactivity disorder localizes pleiotropic loci. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 46,* 1045-1056.

Gjone, H., Stevenson, J., & Sundet, J. M. (1996). Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 588-596.

Gould, M. S., Bird, H., & Staghezza Jaramillo, B. (1993). Correspondence between statistically derived behavior problem syndromes and child psychiatric diagnoses in a community sample. *Journal of Abnormal Child Psychology, 21,* 287-313.

Hebebrand, J., Dempfle, A., Saar, K., Thiele, H., Herpertz-Dahlmann, B., Linder, M. et al. (2006). A genome-wide scan for attention-deficit/hyperactivity disorder in 155 german sib-pairs. *Molecular Psychiatry, 11,* 196-205.

Hudziak, J. J. (2001). The role of phenotypes (diagnoses) in genetic studies of attention-deficit/hyperactivity disorder and related child psychopathology. *Child and Adolescent Psychiatric Clinics of North America, 10,* 279-97, viii.

Hudziak, J. J., Copeland, W., Stanger, C., & Wadsworth, M. (2004). Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45,* 1299-1307.

Hudziak, J. J., Rudiger, L. P., Neale, M. C., Heath, A. C., & Todd, R. D. (2000). A twin study of inattentive, aggressive and anxious/depressed behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 469-476.

Kuntsi, J. & Stevenson, J. (2001). Psychological mechanisms in hyperactivity: II The role of genetic factors. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 211-219.

Lengua, L. J., Sadowski, C. A., Friedrich, W. N., & Fisher, J. (2001). Rationally and empirically derived dimensions of children's symptomatology: expert ratings and confirmatory factor analyses of the CBCL. *Journal of Consulting and Clinical Psychology, 69,* 683-698.

Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data*. (2 ed.) New York: Wiley and Sons.

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

Martin, N., Scourfield, J., & McGuffin, P. (2002). Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. *British Journal of Psychiatry, 180,* 260-265.

Nadder, T. S. & Silberg, J. L. (2001). Comparison of multiple measures of ADHD symptomatology: a multivariate genetic analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42,* 475-486.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA: Department of Psychiatry.

Ogdie, M. N., MacPhie, I. L., Minassian, S. L., Yang, M., Fisher, S. E., Francks, C. et al. (2003). A genomewide scan for attention-deficit/hyperactivity disorder in an extended sample: Suggestive linkage on 17p11. *American Journal of Human Genetics, 72,* 1268-1279.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral Genetics*. New York.

Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2003). Heritability of attention problems in children: I. Cross-sectional results from a study of twins, age 3-12 years. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 117B,* 102-113.

Rietveld, M. J. H., Valk van der, J. C., Bongers, I. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D. I. (2000). Classification of twin zygosity. *Twin Research, 3,* 134-141.

Shaffer, D., Schwabstone, M., Fisher, P., Cohen, P., Piacentini, J., Davies, M. et al. (1993). The Diagnostic Interview Schedule for Children Revised Version (Disc-R) .1. Preparation, Field Testing, Interrater Reliability, and Acceptability. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 643-650.

Sherman, D. K., McGue, M. K., & Iacono, W. G. (1997). Twin concordance for attention deficit hyperactivity disorder: A comparison of teachers' and mothers' reports. *American Journal of Psychiatry, 154,* 532-535.

Simonoff, E., Pickles, A., Hervas, A., Silberg, J. L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine, 28,* 825-837.

Sprafkin, J., Gadow, K. D., Salisbury, H., Schneider, J., & Loney, J. (2002). Further evidence of reliability and validity of the Child Symptom Inventory-4: Parent Checklist in clinically referred boys. *Journal of Clinical Child and Adolescent Psychology, 31,* 513-524.

Todd, R. D., Constantino, J. N., & Neuman, R. J. (2005). Redefining early-onset disorders for genetic studies: attention -deficit/hyperactiviy disorder and autism. In C.F.Zorumski & E. H. Rubin (Eds.), *Psychopathology in the genome and neuroscience era* (1 ed., pp. 33-49). Arlington: American Psychiatric Publishing, Inc.

Verhulst, F. C., Ende van der, J., & Koot, H. M. (1996). *Handleiding voor de CBCL/4-18*. Rotterdam: Afdeling Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.

Vierikko, E., Pulkkinen, L., Kaprio, J., & Rose, R. J. (2004). Genetic and environmental influences on the relationship between aggression and hyperactivity-impulsivity as reated by teachers and parents. *Twin Research, 7,* 261-274.

CHAPTER

12

ATTENTION PROBLEMS AND ATTENTION DEFICIT HYPERACTIVITY DISORDER IN DISCORDANT AND CONCORDANT MZ TWINS: EVIDENCE OF ENVIRONMENTAL MEDIATORS

This article is in press as Eske M. Derks, Hanne Lehn, James J Hudziak, Peter Heutink, Toos C. E. M. van Beijsterveldt, and Dorret I. Boomsma. Attention problems and attention deficit hyperactivity disorder in discordant and concordant MZ twins: Evidence of environmental mediators. *Journal of the American Academy of Child and Adolescent Psychiatry, in press.*

ABSTRACT

The aim of this study is to investigate environmental influences on attention problems (AP) and attention deficit hyperactivity disorder (ADHD) in Monozygotic (MZ) twins discordant and concordant high and low for these traits. Ninety-five twin pairs from the Netherlands Twin Register who were followed longitudinally since birth were selected. Survey data were collected at ages 1, 2, 3, 5, 7, 10 and 12 years from mothers, fathers, twins and teachers. The mother participated in a structured clinical interview when twins were between 10 and 17 years old. Affected twins from discordant pairs scored higher than unaffected co-twins on multiple measures of AP, ADHD and other behavior problems according to mother, teacher and self. Behavioral discordance was evident at age 2 and all subsequent measurements. Compared to unaffected co-twins, affected twins had lower birth weight and delayed physical growth and motor development. Differences between discordant and concordant groups were reported for maternal smoking, sleeping in different rooms, and living with only one parent. Significant markers of ADHD are found in infancy and include low birth weight and delayed motor development. As the knowledge of specific genetic and environmental influences on ADHD increases, future studies may focus on their complex interplay.

INTRODUCTION

In the field of developmental psychopathology there is increasing enthusiasm for the study of specific genetic and environmental influences and their interactions on Attention Deficit Hyperactivity Disorder (ADHD) (Kahn et al., 2003). There are a variety of approaches which may be used to estimate and identify these genetic and environmental components and their interactions. This report presents a powerful method to identify environmental influences on ADHD, namely the monozygotic discordant twin design.

The etiology of ADHD has been studied extensively over the past two decades. Twin, family, and adoption studies all provide evidence for significant genetic influences on ADHD (Derks et al., in press; Faraone & Doyle, 2000). Based on statistical modeling of a variety of taxonomic measurements of ADHD symptom domains in twin samples, heritability estimates of at least 60% have been obtained. This implies that evironmental factors may explain as much as 40% of the etiology of ADHD. The search for environmental factors that contribute to the development of ADHD has yielded a number of candidates . These include poor parenting strategies and family dysfunction (Biederman et al., 2002), low parental socio-economic status (SES) and environmental deprivation (Ornoy, 2003), food additives, (Boris & Mandel, 1994), maternal smoking (Thapar et al., 2003), maternal alcohol consumption (Knopik et al., 2005), and traumatic brain injury (Bloom et al., 2001). The strongest evidence concerns insults that occur during the pre- and peri- neonatal period such as intrauterine exposure to nicotine which has repeatedly been associated with increased (up to twofold) risk of ADHD (reviewed in Linnet et al., 2003). This relation between maternal smoking during pregnancy and ADHD in offspring remains significant after controlling for parental ADHD status (e.g., Mick et al., 2002). Children who are born prematurely and with low birth weight are also at increased risk of developing symptoms of ADHD (for meta-analysis, see Bhutta et al., 2002). The relative value of adversity, low birth weight, exposure to intrauterine alcohol, cocaine, nicotine, lead, and viral infections have all been hypothesized as potential contributors to the etiopathology of ADHD.

As molecular genetic techniques have improved remarkably over the past two decades it is now possible to combine the study of environmental mediators with specific gene finding expeditions (candidate gene studies). For example, Kahn et al. (2003) report that prenatal nicotine exposure and a particular DAT polymorphism (DAT 10 Repeat)

increase the risk of ADHD only when both risk factors are present. Early findings such as these have increased the need to refine both our molecular genetic and environmental assessment strategies.

The field of behavioral genetics has long considered modifications of the classic twin design to account for more than just identifying the magnitude of environmental effects. One such unique application of the twin design, and the focus of this report, is the investigation of specific, unique, environmental influences on ADHD through the Monozygotic Twin Difference Method (MZD) (Martin et al., 1997). Because MZ twins nearly always have identical genomes, most differences in their behavior must be due to the effects of environmental influences, which may act directly on the phenotype, or, for example, through post-genomic modifications via methylation processes (Fraga et al., 2005). A group of MZ twins discordant for ADHD has previously been described in the literature in terms of clinical characteristics (Sharp et al., 2003) and brain anatomy (Castellanos et al., 2003). These MZ discordant twins demonstrated decreased familiality of ADHD in terms of lower symptom scores in fathers when compared with affected singletons. Also, affected twins had lower birth weight, were more likely to present in breech position, and had volumetric reductions in caudate nucleus compared to unaffected co-twins.

This study expands upon this prior work by using prospective data to evaluate and test differences within MZ discordant pairs. We also compare the MZ discordant pairs to MZ concordant pairs. We look at a large range of environmental mediators such as maternal smoking and alcohol use during pregnancy, duration of pregnancy, placenta sharing, birth weight and height, time in the incubator and medical complications. In addition developmental processes are considered such as rate of maturation, physical health, and medical histories. Environmental factors such as sharing a bedroom or class room, and living with only one parent are also examined.

The aim of this study is to identify and describe environmental mediators of AP and ADHD. MZ twins were selected for discordance in AP symptom scores on the Child Behavior Checklist (CBCL) and ADHD diagnosed from the Diagnostic Interview Schedule for Children (DISC). Two groups of MZ concordant pairs were included: concordant for high AP/ADHD and concordant for low AP/ADHD. All twin pairs were recruited from the Netherlands Twin Register, which consists of over 25,000 twin pairs followed prospectively since birth. Their parents and teachers participate in survey studies and provide the data for prospective analyses of AP/ADHD environmental risk factors.

METHOD

SUBJECTS

All data originate from the Netherlands Twin Register (NTR; Boomsma et al., 2002). Twins enroll in the longitudinal survey studies of the NTR, which focus on growth, health and the development of twins' behavior and behavior problems. Surveys are sent to the parents when the twins are 1, 2, 3, 5, 7, 10 and 12 years old and to the teachers from age 7 onwards.

Ninety-five MZ twin pairs participated in this study. These pairs were selected from two ongoing studies of the NTR. One study (henceforth referred to as "Wave I") combines information from multiple informants, time points and assessment techniques in order to identify heritable phenotypes for ADHD. The other study ("Wave II") uses magnetic resonance imaging (MRI) to trace symptoms of ADHD back to abnormalities in neural structure and processing. The selection procedures employed in Waves I and II are summarized in Figure 1.

For Wave I, children were selected from the birth cohorts 1989-1992, and for Wave II from the birth cohorts 1986-1994. For both Waves, subjects who were likely to be MZ twins were selected among twins whose mothers had completed the NTR surveys at ages 7, 10

and/or 12 years at least at two time points. Within the remaining sample, discordant and concordant twin pairs were selected for interview participation. A twin pair was initially regarded as discordant if one twin scored high on AP measured with the CBCL (T 60 at all available time points and T 65 at least once). A twin pair was regarded concordant low if both twins had low AP scores (T 55 at all available time points) and concordant high if both twins had high scores. Discordant and concordant pairs were matched on gender, zygosity, date of birth, maternal age, and parental SES. For Wave II, matching criteria were gender, date of birth, zygosity, handedness and SES. Twins with severe physical or mental disabilities (e.g., autism, blindness) were excluded. No other Axis I, II or III disorders were excluded.

A total of 183 twin-pairs were selected for participation (128 in Wave I and 55 in Wave II). Of these families, 158 pairs (86%) successfully participated. For the present study, twin-pairs were only selected in Wave II if they were not already selected in Wave I.

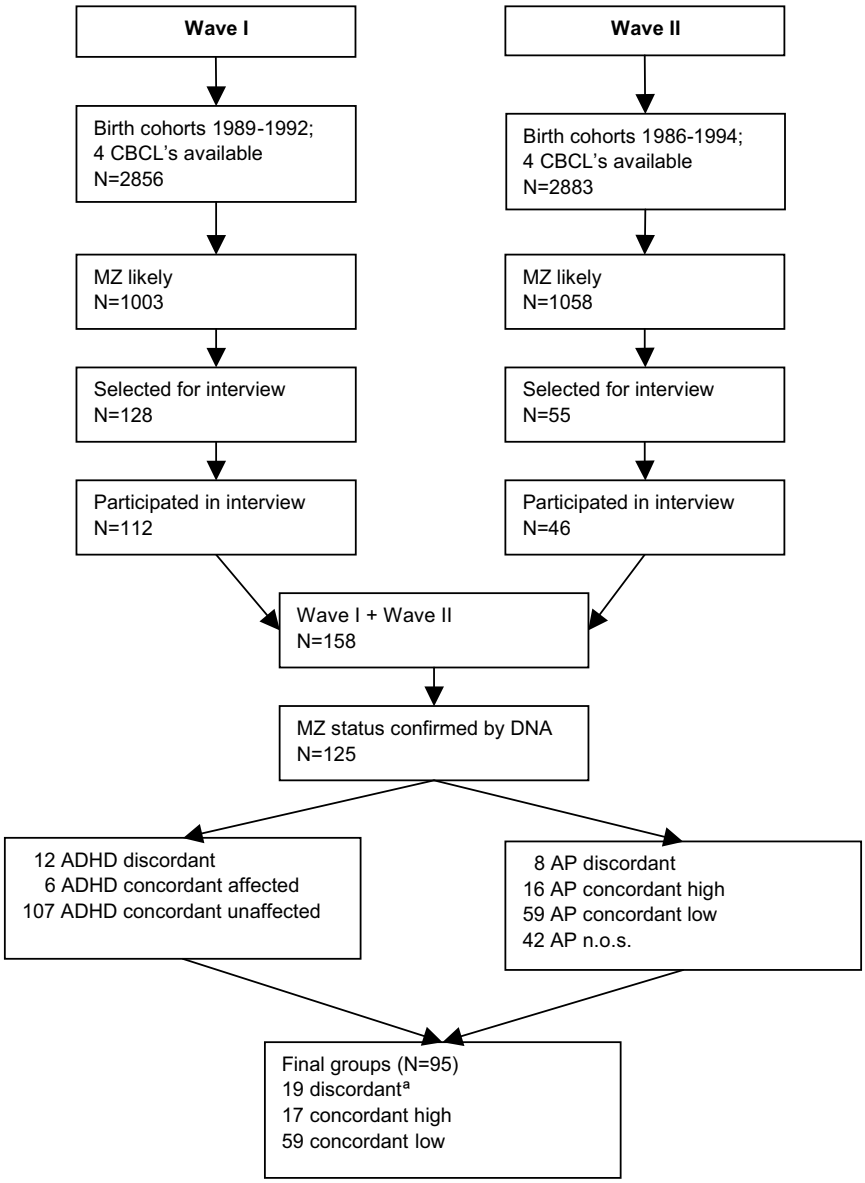
Of the group of 158 families, 120 mothers, 104 fathers and 135 twin pairs returned a DNA sample. Currently, zygosity testing has been completed for 132 pairs. Among these pairs, DNA testing revealed that 125 pairs were indeed MZ. The remaining 7 pairs were DZ and were excluded from statistical analyses.

The mothers of the twins were interviewed with the DISC (see below), and the diagnostic data were used to determine the twins' ADHD status (affected/unaffected; based on type A criteria of the DSM-IV: 6 symptoms/< 6 symptoms). By combining ADHD status with AP status (high/low) derived from the CBCL, three groups were defined: *I) Discordant MZ twins* (ADHD discordant and/or AP discordant; n = 19 pairs); *II) Concordant-high MZ twins (*ADHD concordant affected insofar not AP discordant, and/or AP concordant high insofar not ADHD discordant; n = 17 pairs); *III) Concordant-low MZ twins* (ADHD concordant unaffected and AP concordant low; n = 59 pairs). This classification excluded 29 MZ twin pairs who were ADHD concordant unaffected and AP not otherwise specified (*n.o.s.*; one twin had AP scores that were neither high nor low). The study procedures conformed to the guidelines of our local ethical committee. Mothers provided written informed consent prior to participation.

MEASURES

Clinical Interview. The mother completed a clinical interview, administered over the phone by trained medical students and tape-recorded in a majority of cases. The interview was based on the Dutch version of the DISC IV Parent Version (DISC-IV-P; Ferdinand & Ende van der, 1998). This is a highly structured interview designed for 6 to 17 year-old children. Diagnostic criteria are based on the type A criteria of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994). In addition to the DISC-IV-P, the interview contained questions about the use of psychotherapeutic drugs and the home environment. The 11 items belonging to the AP subscale of the CBCL were also administered.

Figure 1. Selection of MZ Twin Pairs Discordant and Concordant for AP/ADHD



Note: [a]One discordant pair was excluded from the analysis because the unaffected twin used Ritalin at the time of the clinical interview. AP = attention problems; N = number of twin pairs; n.o.s. = not otherwise specified. The selection criteria can be found in text.

Post Interview Questionnaires to Mothers and Twins.

After the clinical interview was administered, mothers completed the Child Behavior Checklist (CBCL/4-18; Achenbach, 1991b) and the Strength and Weaknesses of ADHD-symptoms and Normal Behavior scale (Swanson et al., 2006). With the mother's permission, all twins aged 11 years or older were invited to complete the Youth Self Report (YSR; Achenbach, 1991a).

DNA zygosity testing.

A few months after the clinical interview, all families who had participated were asked to provide buccal samples for DNA extraction. Zygosity testing included Polymerase Chain Reaction (PCR) of eleven highly polymorphic genetic markers.

Longitudinal NTR surveys.

Prospective data were obtained from the longitudinal NTR surveys. The age 1 survey was completed within a few months of the twins' birth. Mothers reported characteristics of the pregnancy, birth, and postnatal period. At age 2, the NTR survey assessed twins' behavioral characteristics, medical history, motor development and physical growthAt age 3, the NTR survey included the CBCL (CBCL/2-3; Achenbach, 1992), in addition to questions about the twins' medical history and physical growth. At age 5, the NTR survey included 42 items from the Devereux Child Behaviour rating scale (DCB; Beijsterveldt van et al., 2004; Spivack & Spotts, 1966). Questions about day care, school, home environment, medical history, growth, and language- and motor development were also included in the survey at age 5.

At ages 7, 10, and 12 the NTR surveys included the CBCL/4-18. Information was also collected about physical and mental disabilities, weight, height, school situation, medical conditions and use of medication or receiving health care.

Parental SES was assessed at ages 7 and 10 years and was based on a full description of the occupation of the parents. The level of occupation was coded according to the system used by Statistics Netherlands (Centraal Bureau voor de Statistiek, 1993) into 5 levels based on the mental complexity of the work and ranged from low skilled (1) to academic work (5). According to Statistics Netherlands, the percentage of families in each of the classes is 7, 25, 40, 20, and 8%, respectively. The mean SES in the general population is 2.97. For this paper, the most recent data available for parental SES were used. When the occupation information was present for both parents, the highest level was taken.

From age 7 onwards, teacher data were collected once permission to approach the teacher was obtained from the parents. The Teacher Report Form (TRF; Achenbach, 1991a) was sent at ages 7, 10 and 12 years. The Attention Problem scale (20 items) derived from TRF at age 12 was analyzed in this paper.

Statistical Analyses

Within Pairs Analyses.

Analyses of child-specific variables (e.g., birth weight) were restricted to discordant pairs. Paired differences (i.e., affected twin vs. unaffected co-twin) were tested for significance with *t-tests for two related samples* (continuous data), *Wilcoxon signed-rank test* (ordinal data), and *McNemar* χ^2*-tests for matched pairs* (nominal data). One-tailed probabilities were used as the affected twins were expected to show more adversity than their unaffected co-twins.

Between Pairs Analyses.

Variables that reflect correspondence between members of a twin pair (e.g., same or different schools) were compared across the three groups of twins. Variables that reflect characteristics that are shared by both members of a twin pair (e.g., maternal smoking) were tested with *one-way ANOVAs* (continuous data), *Kruskal-Wallis tests* (ordinal data), and χ^2*-tests* (nominal data). Post-hoc comparisons included the χ^2*-test* (nominal data) and the *Mann-Whitney U test* (ordinal data).

Two-tailed probabilities were used as we had no clear expectation regarding the direction of effects. Because of the small sample size, alpha was set at .10 to ensure sufficient statistical power and to minimize the risk of making Type 2 errors.

RESULTS

General characteristics

We identified 19 MZ twin pairs discordant for attention problems (Table 1). This group consists of 9 male and 10 female pairs who where between 11 and 15 years old (m = 13.36, sd = 1.54) at the time of the clinical interview. Eight pairs are AP discordant and 11 pairs are ADHD discordant. Of the pairs who met criteria for ADHD, none hade received prior treatment except in one family.The discordant twins had few non-psychiatric medical conditions. However, in one set of twins (no. 3) the affected twin was developmentally retarded and suffered from dyspraxia. Both twins from pair no. 6 have diabetes; and both members of pair no. 15 were born with craniofacial anomalies (affected twin: cleft lip; unaffected co-twin: cleft lip, jaw and palate). Parental SES is rated 4 in a majority (58%) of the MZ discordant twin pairs, with the remaining pairs rated as 2 (21%) or 3 (21%). Average maternal age at time of birth was 30.05 years (sd = 4.00), which is similar to the average age of 29.97 in the NTR 1986-1992 cohorts.

There were 17 MZ twin pairs concordant-high for attention problems (7 male and 10 female pairs; mean age = 13.07 years, sd = 1.98). The female majority in the concordant affected pairs may seem unusual, but it should be realized that the selection was based on T-scores. Because these T-scores were computed separately in boys and girls, the number of affected boys and girls is similar.

Table 1. Demographics and ADHD Symptom Classification of MZ Twins Discordant for AP/ ADHD

Twin pair no.	Age (years)[A]	Sex	AP scores affected/ unaffecte	ADHD affected/ unaffected	ADHD subtype affected
1	15.77	M	high / low	no / no	-
2	15.41	F	high / low	no / no	-
3	15.37	M	high / low	no / no	-
4	15.05	M	high / low	no / no	-
5	14.88	F	high / low	no / no	-
6	14.98	M	high / high	yes / no	IA
7	13.96	F	high / low	no / no	-
8	12.99	F	~ / ~	yes / no	H/I
9	13.00	F	~ / ~	yes / no	IA
10	13.94	M	high / high	yes / no	C
11	12.95	F	high / high	yes / no	IA
12	12.08	M	high / ~	yes / no	H/I
13	12.01	F	high / ~	yes / no	H/I
14	11.58	M	high / high	yes / no	IA
15	11.52	M	high / low	no / no	-
16	11.38	F	~ / ~	yes / no	IA
17	12.79	F	high / low	no / no	-
18	11.03	F	high / ~	yes / no	H/I
19	13.08	M	high / ~	yes / no	IA

Note:[A]Age at the time of the clinical interview. ADHD = Attention Deficit Hyperactivity Disorder; MZ = monozygotic; AP = attention problems measured with the Child Behavior Checklist; M = male; F = female; AP scores high = T-score 60 at all available time points and 65 at least once;AP scores low = T-score 55 at all available timepoints; AP scores ~ = T-scores fall in neither of the statuses high and low; ADHD yes = twin meets DSM-IV Type A criteria; ADHD no = twin does not meet DSM-IV Type A criteria; IA = inattentive subtype;H/I = hyperactive/ impulsive subtype; C = combined subtype.

The group of concordant-high twin pairs includes 11 pairs that are AP concordant, 5 pairs that are ADHD concordant, and one pair that is both AP- and ADHD concordant. Of the 34 concordant-high children, 12 children met criteria for ADHD. Four children (both twins in one pair and the first born twin in two pairs) used Ritalin at the time of the interview. Of the concordant-high twin pairs, 71% has a parental SES of 3, with the remaining pairs rated as 1 (6%), 4 (18%) or 5 (6%). Maternal age at time of birth is 28.16 years on average (sd = 3.45).

There were 59 MZ twin pairs concordant-low for attention problems (30 male and 29 female; age m = 13.01 years, sd = 1.41). In this group, most twins have a parental SES of either 3 (37%) or 4 (36%), with the remaining pairs rated as 2 (7%) or 5 (20%). Maternal age is 30.06 years on average (sd = 3.66).

The three groups of MZ twins did not differ significantly in terms of age, sex or maternal age. A significant group difference was found in parental SES (*Kruskal-Wallis* χ^2 = 5.00; $p_{\text{two-tailed}}$ = .082). Parental SES is significantly higher among concordant-low twins than among concordant-high twins (*Mann-Whitney U* = 342.50; $p_{\text{two-tailed}}$ = .034). The parental SES of discordant twins lies between that of the two other groups, but differs significantly from neither.

Current Behavior Problems

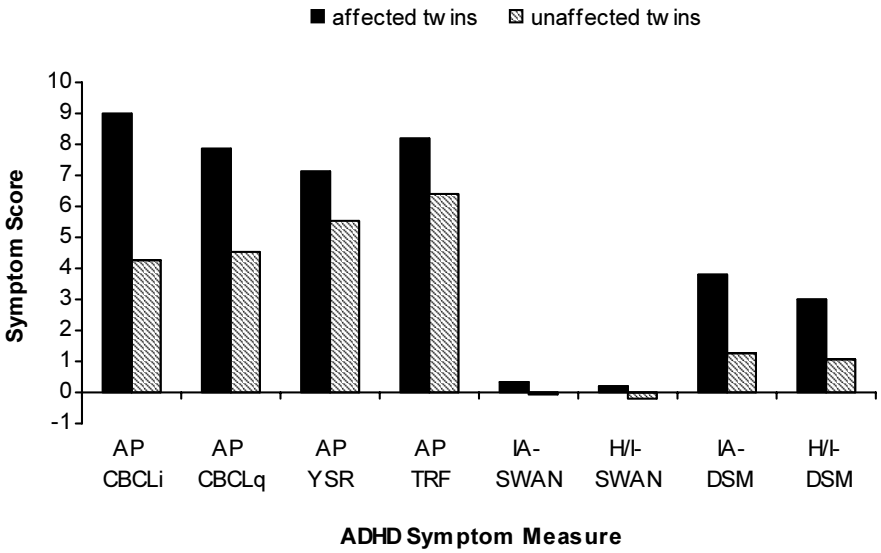
Affected members of the MZ discordant twin pairs score higher than their unaffected co-twins on all measures of ADHD-related symptoms (Figure 2).The difference within discordant pairs in ADHD-related symptoms is paralleled by differences in multiple other problem behaviors. This is most evident in the self-report data.

Compared to their unaffected co-twins, affected twins report more rule breaking behavior (m_{aff} = 3.40, sd_{aff} = 1.30; m_{unaff} = 2.80, sd_{unaff} = 1.74; $t(14)$ = 2.07, $p_{\text{one-tailed}}$ = .029), have more social problems (m_{aff} = 3.67, sd_{aff} = 2.72; m_{unaff} = 2.47, sd_{unaff} = 2.07; $t(14)$ = 1.75, $p_{\text{one-tailed}}$ = .051), and are more aggressive (m_{aff} = 10.93, sd_{aff} = 4.54; m_{unaff} = 7.67, sd_{unaff} = 4.45; $t(14)$ = 2.53, $p_{\text{one-tailed}}$ = .012), withdrawn (m_{aff} = 3.40, sd_{aff} = 2.69; m_{unaff} = 2.33, sd_{unaff} = 2.53; $t(14)$ = 2.62, $p_{\text{one-tailed}}$ = .010), and anxious/depressed (m_{aff} = 8.20, sd_{aff} = 5.51; m_{unaff} = 5.27, sd_{unaff} = 4.62; $t(14)$ = 2.62, $p_{\text{one-tailed}}$ = .073). The maternal CBCL-ratings yield significant differences on aggressive behavior (m_{aff} = 11.41, sd_{aff} = 7.64; m_{unaff} = 7.76, sd_{unaff} = 7.21; $t(16)$ = 2.29, $p_{\text{one-tailed}}$ = .018) and social problems (m_{aff} = 3.06, sd_{aff} = 3.17; m_{unaff} = 1.82, sd_{unaff} = 1.94; $t(16)$ = 2.11, $p_{\text{one-tailed}}$ = .025). Psychiatric diagnoses other than ADHD are also found exclusively in affected children; in pair no. 12 (social phobia and oppositional disorder) and in pair no. 19 (specific phobia for blood and wounds).

Development

The longitudinal data show that the differences in ADHD symptomatology were present at an early age and were relatively stable throughout development (Figure 3). Also with regard to other behavior problems, differences within discordant pairs have been present since toddlerhood. Figure 4 shows the differences in aggressive behavior and social problems, as these are most prominent and also most persistent across time.The earliest reports of the twins' behavioral characteristics were collected at age 2 and concern sleep, crying, and amount of distress from disturbances in the daily routine. During the first 18 months of life, affected twins cried significantly more often (Wilcoxon Z = -1.41, $p_{\text{one-tailed}}$ = .079) than unaffected co-twins.

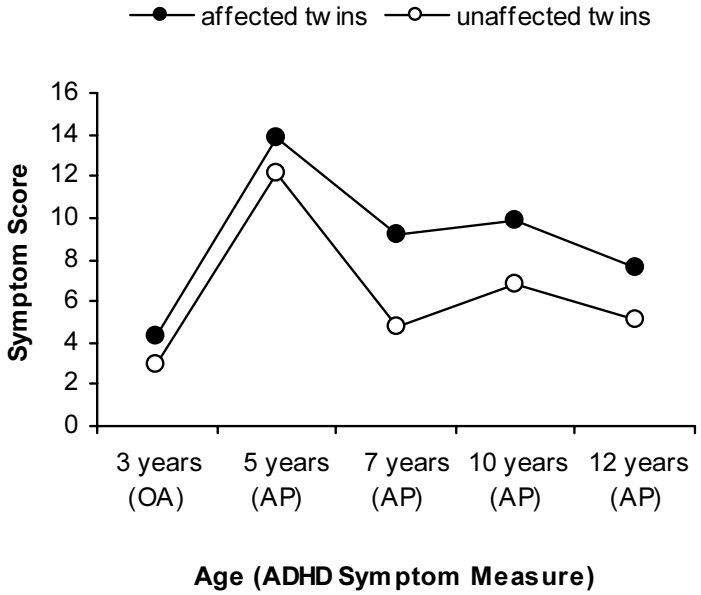
Figure 2. ADHD Symptom Ratings in Monozygotic Twins Discordant for AP/ADHD



Note: AP = attention problems measured with the Child Behavior Checklist; IA = symptoms of inattention; H/I = symptoms of hyperactivity/impulsivity; CBCL = Child Behavior Checklist; YSR = Youth Self Report; SWAN = Strengths and Weaknesses of ADHD-symptoms and Normal Behavior scale; DSM = Diagnostic and Statistic Manual of Mental Disorder; i = ratings obtained during the clinical interview; q = ratings obtained with questionnaire after the clinical interview; * = $p < .10$; ** = $p < .05$; *** = $p < .01$; **** = $p > .005$; ***** = $p < .001$. Scales differ in range (AP CBCLi, AP CBCLq, AP YSR and AP TRF: 0-22; IA-SWAN and H/I-SWAN: -3-3; IA-DSM and H/I-DSM: 0-9). Figure is based on data from 19, 17, 15, 11, 16, 16, 19, and 19 (for symptom measures left to right) complete twin pairs.

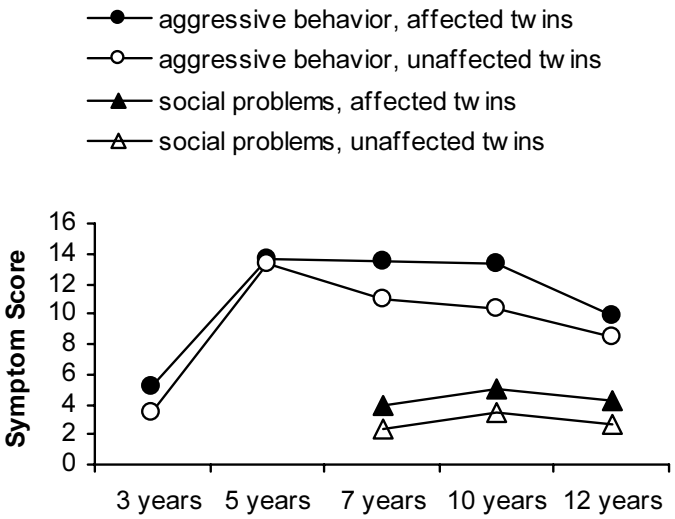
The affected twins also lagged behind their unaffected co-twins in early motor development. They were slightly but significantly older when they learned how to roll over (m_{aff} = 6.88 months, sd_{aff} = 2.26; m_{unaff} = 6.38 months, sd_{unaff} = 1.65; $t(15) = 1.44$, $p_{one\text{-}tailed} = .086$) and to sit upright ($m_{aff}$ = 9.47 months, sd_{aff} = 2.97; m_{unaff} = 8.89, sd_{unaff} = 2.23; $t(17) = 2.08$, $p_{one\text{-}tailed} = .027$). The passing of later milestones in motor development was not significantly different within discordant pairs.

Figure 3. Development of ADHD Symptoms in Monozygotic Twins Discordant for AP/ADHD



Note: OA = overactivity measured with the Child Behavior Checklist; AP = attention problems measured with Devereux Child Behaviour rating scale at age 5 and with the Child Behavior Checklist at ages 7, 10, and 12; * = p < .10; ** = p < .05; *** = p < .01; **** = p > .005. Scales differ in range (OA age 3: 0-10; AP age 5: 0-25; AP ages 7, 10 and 12: 0-22). Figure is based on data from 15, 17, 17, 18, and 15 (for ages left to right) complete twin pairs.

Figure 4. Development of Additional Behavior Problems in Monozygotic Twins Discordant for AP/ADHD



Note: AP = attention problems; * = p < .10; ** = p < .05; *** = p < .01; **** = p > .005. Scales differ in range (aggressive behavior age 3: 0-18; aggressive behavior age 5: 0-35; aggressive behavior ages 7, 10 and 12: 0-40; social problems ages 7, 10, and 12: 0-16). Social problems were not assessed at ages 3 and 5). Figure is based on data from 15, 17, 17, 18, and 15 (for ages left to right) complete twin pairs.

We found evidence of a growth discrepancy between affected and unaffected twins during the first two years of life. Affected twins weighed less than their unaffected co-twins around ages 6 months (m_{aff} = 6589.44 g, sd_{aff} = 1126.76; m_{unaff} = 6910.28 g, sd_{unaff} = 837.02; $t(17)$ = -1.74, $p_{one\text{-}tailed}$ = .050), 1 year (m_{aff} = 9099.72 g, sd_{aff} = 1539.82; m_{unaff} = 9375.56 g, sd_{unaff} = 1226.16; $t(17)$ = -1.38, $p_{one\text{-}tailed}$ = .093), and 2 years (m_{aff} = 11967.69 g, sd_{aff} = 2690.85; m_{unaff} = 12536.92 g, sd_{unaff} = 2493.12; $t(12)$ = -1.65, $p_{one\text{-}tailed}$ = .062). Similarly, affected twins were of shorter statue at 6 months (m_{aff} = 64.33 cm, sd_{aff} = 4.28; m_{unaff} = 65.14 g, sd_{unaff} = 3.04; $t(15)$ = -1.61, $p_{one\text{-}tailed}$ = .065), 1 year (m_{aff} = 74.38 cm, sd_{aff} = 3.71; m_{unaff} = 75.04 cm, sd_{unaff} = 3.34; $t(15)$ = -1.41, $p_{one\text{-}tailed}$ = .089), and 2 years (m_{aff} = 86.54 cm, sd_{aff} = 6.33; m_{unaff} = 87.96 cm, sd_{unaff} = 5.76; $t(12)$ = -2.03, p = .033). At age 3, neither weight (m_{aff} = 13970.00 g, sd_{aff} = 2802.40; m_{unaff} = 14290.00 g, sd_{unaff} = 2147.58) nor height (m_{aff} = 96.05 cm, sd_{aff} = 5.79; m_{unaff} = 96.00 cm, sd_{unaff} = 5.27) differed within discordant pairs.

Further characteristics of discordant twins

In discordant pairs, affected twins were about equally often the first born (n = 9 pairs) as the second born (n = 10 pairs). The average birth weight of affected twins was significantly lower (m = 2421.84, sd = 542.23) than the average birth weight of unaffected co-twins (m = 2591.05, sd = 517.22): $t(18)$ = -1.99, $p_{one\text{-}tailed}$ = .031. No difference was found in the twins' height at birth.

The incidence of neonatal medical complications was similar in affected and unaffected twins. However, for those cases where both twins were in the incubator (n = 12 pairs), the duration of treatment was significantly longer for the affected (m_{aff} = 12.51 days, sd_{aff} = 16.02) than for the unaffected (m_{unaff} = 8.84 days, sd_{unaff} = 13.66) twin ($t(11)$ = 1.62, $p_{one\text{-}tailed}$ = .067).

Except for the initial differences in maturation, both members of most discordant pairs have been in good physical health. Their medical histories are dominated by harmless injuries (e.g., cuts and wounds) and common diseases of childhood (e.g., ear infections and chickenpox). The only severe condition reported is the occurrence of meningitis at age five in one of the affected twins (pair no. 5). Within pairs, there are no significant differences in the number of general practitioner consults, hospitalization, anaesthetization, or drug treatments.

Environmental Mediators of Behavioral Discordance

A significant difference was found in the frequency with which mothers in the discordant, concordant high and concordant low groups reported smoking during pregnancy ($\chi^2(2)$ = 6.25, $p_{two\text{-}tailed}$ = .044). Maternal smoking was more common in the group of concordant-high twins (n = 6/17, 35%) than in the group of concordant-low twins (n = 6/59, 10%): $\chi^2(1)$ = 6.27, $p_{two\text{-}tailed}$ = .012). Discordant twins (maternal smoking reported in n = 4/19, 21% of pairs) did not differ from the concordant groups on this measure. No group differences were found in other pregnancy characteristics; i.e. pregnancy duration, maternal alcohol consumption, and proportion of twins sharing placenta.

During the first two years of life, discordant twins slept in the same room in a majority of cases (n = 17/18, 94%). A similar pattern is found among concordant-high twins (n = 17/17, 100%) and concordant-low twins (n = 54/57, 95%), and there is no significant difference across groups. Sleeping in separate rooms becomes more common at later ages, and significant group differences appear both at age 2-4 years ($\chi^2(2)$ = 5.16, $p_{two\text{-}tailed}$ = .076) and age 4-6 years ($\chi^2(2)$ = 4.79, $p_{two\text{-}tailed}$ = .091). Sleeping in separate rooms was significantly more common in discordant pairs than in concordant-low pairs both at age 2-4 years (n_{disc} = 4/18, 22%; $n_{conc\text{-}low}$ = 3/57, 5%; $\chi^2(1)$ = 4.65, $p_{two\text{-}tailed}$ = .031 and age 4-6 years (n_{disc} = 6/17, 35%; $n_{conc\text{-}low}$ = 7/57, 12%; $\chi^2(1)$ = 4.79, $p_{two\text{-}tailed}$ = .029). No group differences were found in the proportion of twins pairs who were in the same class and/or at the same school at age 5. A significant difference was found in the proportion of twins who lived

with both parents at age 5 ($\chi^2(2) = 9.10$, $p_{\text{two-tailed}} = .011$). Post hoc analyses revealed that living with only one parent was significantly more common among concordant-high twins ($n_{\text{conc-low}} = 3/17$, 18%) than among discordant twins ($n_{\text{disc}} = 0/18$, 0%; $\chi^2(1) = 3.47$, $p_{\text{two-tailed}} = .062$) or concordant-low twins ($n_{\text{conc-low}} = 1/58$, 2%; $\chi^2(1) = 6.60$, $p_{\text{two-tailed}} = .010$).

DISCUSSION

One of the promises of the genomic era in medicine is that through the discovery of genes that are involved in the etiopathology of common diseases, such as ADHD, advances will be made regarding diagnosis, prognosis, and treatment. This promise is based on a complicated premise, in which identifying one of many risk genes for an illness may lead to a breakthrough in understanding complex illnesses such as ADHD. Although multiple genes of small effect have been identified, principally through candidate gene approaches, none have yielded the kind of information that has led to changes in the way this disorder is assessed or treated. As intoxicating as the promise of the new genomics is, an equally important approach to developing new assessment and treatment approaches, is the study of environmental mediators of complex illness. Here we have tested if environmental contributors either put children at risk for or protect them from developing ADHD.

We chose a study population of MZ twins discordant, concordant-high and concordant-low for AP/ADHD in order to evaluate and identify the contributions of environmental mediators. Our results show that affected children from discordant pairs have higher symptom ratings than unaffected co-twins on multiple measures (CBCL, YSR, TRF, SWAN, DSM/DISC-P) of inattention, hyperactivity and impulsivity, as well as on other behavior problems, including social problems, rule breaking behavior, aggression, withdrawn behavior, and anxiety/depression. Because MZ discordant twins have identical genomes, the differences between the affected and unaffected members are the result of the environment. Our findings are in accordance with Sharp et al. (2003) who also reported cross-rater agreement in symptom severity and substantial comorbidity in a sample of MZ twins discordant for ADHD. The longitudinal data in our study showed that the different behavioral profiles of MZ discordant twins are of early onset and relatively consistent across time. In infancy the affected twins already showed signs of greater behavioral imbalance (e.g. frequent crying). At age 3 they were rated as more overactive. Attention problems were more prominent in affected twins at all ages, whereas persistent symptoms of aggressive behavior and social problems emerged at ages 3 and 7, respectively.

At the core of our analyses were attempts to identify conditions predictive of current AP/ADHD status. In MZ discordant pairs, affected twins experienced more adversity in infancy (lower birth weight and spended more time in the incubator) and were disadvantaged in terms of maturation (delayed physical growth and slower acquisition of motor skills). Low birth weight and prematurity are among the most frequently suggested risk factors for ADHD (Bhutta et al., 2002). Our results do not only back up these suggestions but also indicate the truly environmental origin of these risk factors. The MZ discordant twins described by Sharp et al. (2003) and Castellanos et al. (2003) also provide evidence for the impact of neonatal stressors, namely low birth weight and breech position, and suggest they have an effect through interfering with the formation of striatal circuits. Similarly, a recent paper reports a significant correlation between MZ differences in birth weight and MZ differences in hyperactivity symptoms at age 7 (Asbury et al., 2006).

The inclusion of MZ concordant twins provided reference values for exposure to shared environmental risk factors and for the degree of differential experiences. Group comparisons revealed that maternal smoking during pregnancy and living with only one parent was more common among MZ twins concordant high for AP/ADHD, although the

former was significantly more common only when compared to concordant low twins. Another finding was that discordant twins more often had separate bedrooms during their preschool years.

Clinical Implications

We used a phenotyping strategy that utilized each of the prevailing paradigms of phenotype definition (DSM-IV structured interviews, parent reports, teacher reports, and self reports). When using a wide variety of assessment techniques it is not atypical to get widely variant results. However, in the MZ discordant twins, the affected member of the twin was rated as higher by all informants, on all measures, for symptoms of inattention, hyperactivity from the ADHD stable of symptoms, and also on rule breaking, social problems, aggression, anxious/depression, and withdrawn behavior. This was even true when self-report data (YSR) were analyzed. These data support that children with identical DNA, can be remarkably and consistently different. In fact, the profile that emerges on the affected children goes well beyond simple ADHD, and includes children who are deviant on attention, aggression, anxious/depression, and rule breaking behavior. . We describe key developmental periods in which symptomatic findings may be evident and predictive. For example, using assessment approaches such as those discussed above, these children are identified as behaviorally different at age three. These differences are stable and persistent, remaining still evident at age 12. These data indicate the predictive utility of early screening. There are however, other markers that discriminate at an early age, but do not persist. These include low birth weight, which has long been a ubiquitous marker of a difficult intrauterine environment. The MZ discordant affected twins weighed significantly less at birth than their unaffected co-twin. This weight disparity disappeared by the third year of life. In this example, chart review of birth weight may have some utility in considering environmental stressors that have direct impact on the development of psychopathology at later ages. Another example of a developmental window is the achievement of developmental landmarks. MZ discordant affected children were behind in the achievement of their early landmarks, but not later landmarks. Interestingly, some other mediators thought to be markers of later developmental psychopathology such as being in the incubator, neonatal complications, and pediatric infections and brain injuries were not useful in predicting AP/ADHD status in our sample, although affected children were significantly longer in the incubator.

This work is a small step towards identifying environmental influences. The MZ discordant design is ideal for considering environmental mediators of complex traits as well as providing an elegant approach for the future studies of epigenetic modifications.

Limitations

One limitation of our study is the sample size, even though the selection of twin pairs was from a large register. This may be seen as inevitable, given the strong genetic influences on attention problems and ADHD. Although we chose a liberal threshold, effects of small or even moderate size could have been overlooked. We also admit that the MZ discordant twins described in our paper may not be representative of the general population of children with attention problems and ADHD. We did however exclude the most atypical cases (i.e., twins with severe physical or mental disabilities) from participation.

We are not able to specifically identify at this point which environmental insults are so robust that they may not require 'genetic predisposition' to put a child at risk for ADHD. However, we have identified a strategy that may allow for a more refined study of maternal health behavior as it relates to child outcomes. For example, maternal smoking was more common in MZ pairs concordant for AP/ADHD than in the other groups, providing at least secondary evidence, that smoking, overlaid on top of a common genetic risk, leads to

increased rates of ADHD in offspring. Finally, we did not use DSM-IV age of onset or impairment criteria for ADHD diagnosis to improve comparability with CBCL ratings for which no such criteria are available. This might limit the generalizability of the results.

REFERENCE LIST

Achenbach, T. M. (1991a). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

American Psychiatric Association (1994). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorder : DSM-IV (4th ed.)*. Washington, DC: American Psychiatric Association.

Asbury, K., Dunn, J. F., & Nomin, R. (2006). Birthweight-discordance and differences in early parenting relate to monozygotic twin differences in behaviour problems and academic achievement at age 7. *Developmental Science, 9,* F22-F31.

Beijsterveldt van, C. E. M., Verhulst, F. C., Molenaar, P. C. M., & Boomsma, D. I. (2004). The genetic basis of problem behavior in 5-year-old Dutch twin pairs. *Behavior Genetics, 34,* 229-242.

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm - A meta-analysis. *Jama-Journal of the American Medical Association, 288,* 728-737.

Biederman, J., Faraone, S. V., & Monuteaux, M. C. (2002). Differential effect of environmental adversity by gender: Rutter's index of adversity in a group of boys and girls with and without ADHD. *American Journal of Psychiatry, 159,* 1556-1562.

Bloom, D. R., Levin, H. S., Ewing-Cobbs, L., Saunders, A. E., Song, J., Fletcher, J. M. et al. (2001). Lifetime and novel psychiatric disorders after pediatric traumatic brain injury. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 572-579.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Boris, M. & Mandel, F. S. (1994). Foods and Additives Are Common Causes of the Attention-Deficit Hyperactive Disorder in Children. *Annals of Allergy, 72,* 462-468.

Castellanos, F. X., Sharp, W. S., Gottesman, R. F., Greenstein, D. K., Giedd, J. N., & Rapoport, J. L. (2003). Anatomic brain abnormalities in monozygotic wins discordant for attention deficit hyperactivity disorder. *American Journal of Psychiatry, 160,* 1693-1696.

Centraal Bureau voor de Statistiek (1993). *Standaard Beroepenclassificatie 1992 [Standard Classification of Occupations 1992].* Heerlen, the Netherlands: Centraal Bureau voor de Statistiek.

Derks, E. M., Hudziak, J. J., and Boomsma, D. I. (in press). Genetics of ADHD, hyperactivity, and attention problems. *Handbook of Behavior Genetics.*

Faraone, S. V. & Doyle, A. E. (2000). Genetic influences on attention deficit hyperactivity disorder. *Current Psychiatry Reports, 2,* 143-146.

Ferdinand, R. F. & Ende van der, J. (1998). *Diagnostic Interview Schedule for Children IV Parent-version*. Rotterdam: Erasmus University Rotterdam, Department of Child and Adolescent Psychiatry.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestart, M. L. et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 10604-10609.

Kahn, R. S., Khoury, J., Nichols, W. C., & Lanphear, B. P. (2003). Role of dopamine transporter genotype and maternal prenatal smoking in childhood hyperactive-impulsive, inattentive, and oppositional behaviors. *Journal of Pediatrics, 143,* 104-110.

Knopik, V. S., Sparrow, E. P., Madden, P. A. F., Bucholz, K. K., Hudziak, J. J., Reich, W. et al. (2005). Contributions of parental alcoholism, prenatal substance exposure, and genetic transmission to child ADHD risk: a female twin study. *Psychological Medicine, 35,* 625-635.

Linnet, K. M., Dalsgaard, S., Obel, C., Wisborg, K., Henriksen, T. B., Rodriguez, A. et al. (2003). Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: Review of the current evidence. *American Journal of Psychiatry, 160,* 1028-1040.

Martin, N., Boomsma, D., & Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genetics, 17,* 387-392.

Mick, E., Biederman, J., Faraone, S. V., sayer, J., & Kleinman, S. (2002). Case-control study of Attention-Deficit Hyperactivity Disorder and maternal smoking, alcohol Use, and drug use during pregnancy. *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 378-385.

Ornoy, A. (2003). The impact of intrauterine exposure versus postnatal environment in neurodevelopmental toxicity: long-term neurobehavioral studies in children at risk for developmental disorders. *Toxicology Letters, 140,* 171-181.

Sharp, W. S., Gottesman, R. F., Greenstein, D. K., Ebens, C. L., Rapoport, J. L., & Castellanos, F. X. (2003). Monozygotic twins discordant for Attention-Deficit/Hyperactivity Disorder: ascertainment and clinical characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry, 42,* 93-97.

Spivack, G. & Spotts, J. (1966). *The Devereux Child Behavior (DCB) Rating Scale*. Devon, PA: The Devereux Foundation.

Thapar, A., Fowler, T., Rice, F., Scourfield, J., Bree van den, M., Thomas, H. et al. (2003). Maternal smoking during pregnancy and Attention Deficit Hyperactivity Disorder symptoms in offspring. *American Journal of Psychiatry, 160,* 1985-1989.

CHAPTER

13

SAMPLE CHARACTERISTICS AND DATA COLLECTION

This chapter presents an overview of the data collection procedures utilized in this thesis. Questionnaire surveys were collected via mail by the Netherlands Twin Register, and diagnostic interviews were completed by phone as part of this thesis project. In the first section, I will present a study of the response rates to the surveys and examine differences between responders and non-responders on the level of problem behavior and social economic status. Next, I will give a description of the selection of participants and the data collection procedures in the interview study. The chapter is concluded with the investigation of the effect of birth cohort on the level of attention problems.

1. RESPONSE RATES AND ATTRITION IN THE NETHERLANDS TWIN REGISTER

The Netherlands Twin Register (NTR) consists of two groups of twins and their families: the young twins and the adult twins (Boomsma et al., 2002). This thesis focused on the young twin sample. The young twins were registered at birth by their parents. The first twins were registered around 1986. They are recruited with the help of a commercial organization (Felicitas B.V.), which visits parents of newborns at home. If, during this home visit, it turns out that more than one child was born (e.g., twins, triplets), the parents receive a brochure and a NTR registration card. If the parents decide to return the signed registration card, the first questionnaire is mailed to the mother. When this questionnaire is returned, the family is registered with the NTR, and subsequent surveys are sent to the parents, when the twins reach the ages of 2, 3, 5, 7, 10, and 12 years.

The NTR collects data from multiple informants. The parental surveys include maternal and paternal questionnaires and are sent at the ages 3, 5, 7, 10, and 12 years. These surveys include the Child Behavior Checklist (Achenbach, 1991b) at the ages 3, 7, 10, and 12 years, the Conners Parent Rating Scale (Conners, 2001) at the ages 7, 10, and 12 years, and the Devereux Child Behaviour Rating Scale (Spivack & Spotts, 1966) at age 5 years. In addition, if parents provide permission, questionnaires are sent to the twin's teachers at the ages 7, 10, and 12 years. The teacher questionnaires include the Teacher Report Form (Achenbach, 1991c), and the Conners Teacher Rating Scale (Conners, 2001). Since 2004, self-reports (Youth Self Report) (Achenbach, 1991a) are sent to the twins at the ages 14 and 16 years. In this thesis, parental and teacher ratings of the behavior of the twins were analyzed. In the next section, an overview of the response rates in the NTR is provided, and it is investigated whether the response rates depend on children's problem behavior scores and on parental Social Economic Status (SES).

Parental SES was assessed at ages 3, 7, and 10 years based on a full description of the occupation of the parents. The level of occupation was coded according to the system used by the CBS (Centraal Bureau voor de Statistiek, 1993), and classified into 5 levels (1 low skilled to 5 academic) based on the complexity of the work. If the level of occupation information of both parents was present, the highest level was taken as indicative of SES. The overview is based on the data that are available on March 2006. Because most of the following chapters in this thesis were written before March 2006, the reported numbers of participants in these chapters was lower. However, the conclusions based on the March 2006 sample are unlikely to differ from those based on the earlier or later samples.

NUMBER OF PARTICIPANTS AND RESPONSE RATES

Figure 1 shows the number of surveys are sent to the parents who participate in the longitudinal study of the NTR. These surveys include paternal and maternal ratings. The number of participating families decreases over time. The decrease in number of parental surveys sent from timepoint 1 to timepoint 2 is due to the fact that families that do not respond to the first survey, are not registered, and therefore do not receive surveys at subsequent ages. Families who do not respond at a later timepoint do receive questionnaires at later ages. In addition, the number of surveys that are sent, decreases over time, because some children have yet to reach the target age. Surveys are available for birth cohorts 1986 to 2005 at age 1, and for birth cohorts 1986 to 1993 at age 12. The responders in Figure 1 are members of families in which at least one of the parents returned the questionnaire; the non-responders are the families in which neither parent returned the questionnaire. It should be noted that the non-responders are not necessarily unwilling to return the survey, an alternative reason for non-participation is that the survey was sent to the wrong adress (e.g., the family moved to a new address without informing the NTR).

The response rate of the teachers is shown in Figure 2. The questionnaires are sent to the teachers on an individual basis, therefore, the numbers in Figure 2 represent the number of individual twins, and not twin-pairs, as was the case in the overview of the parental response rate (Figure 1). Teacher ratings are available for birth cohorts 1992 to 1998, 1989 to 1995, and 1986 to 1993 at the ages 7, 10, and 12 years, respectively. The Teacher Report Form was first collected in 1999. Data collection in 7, 10, and 12 years old started at the same time. The birth cohort data show that there is little overlap between the teacher data at the ages 7 and 12 years, which is the reason that no longitudinal analyses have yet been performed using the teacher data. Table 1 shows the total number of questionnaires that are available in the NTR, and the response rates of parents and teachers.

Figure 1. The number of families who received a survey at age 1 until age 12 years

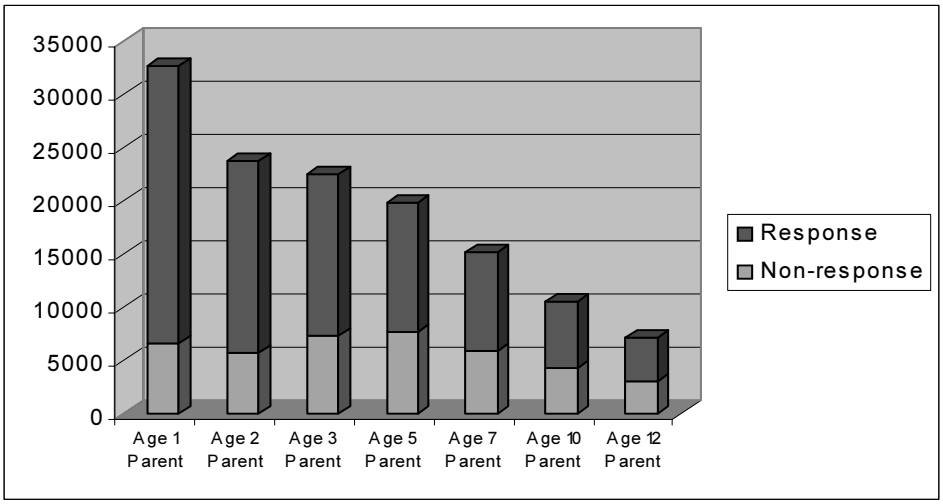


Figure 2. The number of teachers who received a questionnaire at age 7 until age 12 years

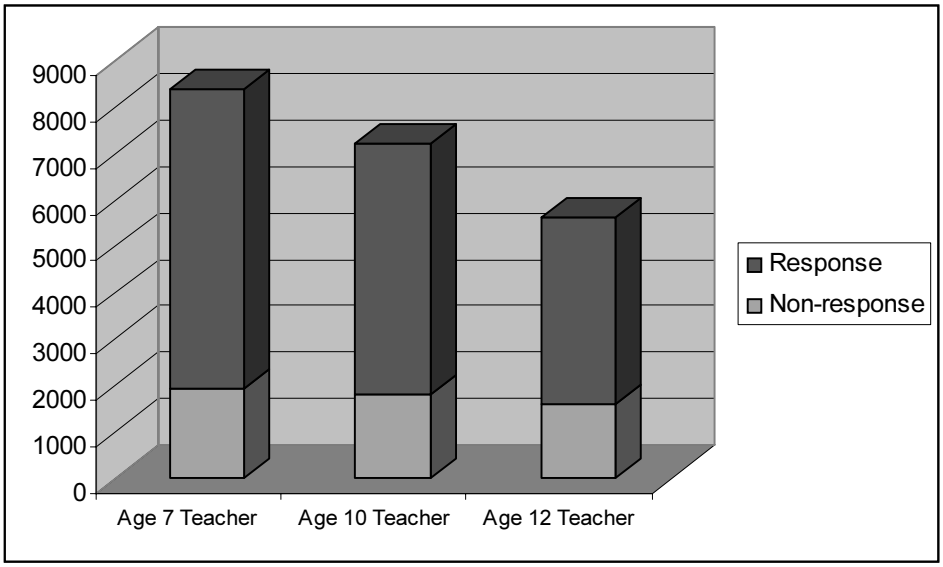


Table 1. Total number of available questionnaires and response rate of parents and teachers

	Parent			Teacher		
	No. surveys returned	Total No. surveys send	Response rate (%)	No. Questionnaires returned	Total No. Questionnaires send	Response rate (%)
Age 1	25991	32613	80	-	-	-
Age 2	18170	23829	76	-	-	-
Age 3	15215	22447	68	-	-	-
Age 5	12169	19775	62	-	-	-
Age 7	9240	15204	61	6451	8379	77
Age 10	6195	10529	59	5425	7235	75
Age 12	4197	7215	58	4023	5633	71

Note: The number of parental surveys represent number of twin-pairs as surveys are sent to the family; the number of teacher questionnaires represent number of individual twins as questionnaires are sent on an individual basis

The response rate was calculated as the number of surveys/questionnaires returned divided by the total number of surveys/questionnaires sent. The response rates of the parents range between 58 and 80%, and show a clear trend of decreasing response rates as children grow older. The response rates of the teachers are quite consistent over time, and vary between 71 and 77%.

Differences between responders and non-responders as a function of previous response, the level of problem behavior and parental SES

With respect to the response rate of the parents, three questions are addressed. First, is the (non)response of a family associated with the (non)response of the family at the previous measurement occasion? Second, is the level of problem behavior associated with the (non)response of the family at the previous age? Third, is the level of Social Economic Status (SES) associated with the (non)response of a family at the previous questionnaire? Appendix 1 shows the results of the analyses that were performed to investigate these questions. Because simultaneous analysis of the data from first- and second born twins may lead to biased test statistics as a result of statistical dependency in the data, the tests are based on the problem behavior scores of the first-born twins only. The (type I error probability) of each test was set on .01.

Response rates at a given age are lower in families that did not respond at the previous measurement occasion (about 25%), compared to families that responded at the previous age (about 75%). These results show that a quarter of the non-responders return to the study at the next target age. With respect to the level of problem behavior at the ages 3, 5, and 12 years, attention problems (AP), aggression (AGG), and anxiety (ANX) scores were not significantly different between families that responded at the previous target age and families that did not respond at the previous target age. However, at age 7, twins obtained higher maternal rated scores on AP, AGG, and ANX in families that did not respond at age 5 than in families that did respond at age 5. At age 10, the only difference was found for AP: twins obtained higher scores when the parents did not respond at age 7 than when the parents did respond. A significant association was also found between (non)response and the level of SES. At ages 7 and 10, the level of SES was higher in families that returned the questionnaire than in families that did not return the questionnaire. In contrast, no difference was found for the level of SES at age 3. Although significant differences in the levels of problem behaviors are found between responders and non-responders, the effect sizes (i.e., proportion of variance explained by the factor) are all .00. This implies that the differences between responders and non-responders are statistically significant, probably as a result of the large sample sizes, but are not practically significant.

In conclusion, the response rate of the parents at a given target age is associated with the response rate at the previous measurement occasion. A relation between problem behavior scores and SES, on the one hand, and response rate, on the other hand, was also found, but appears to be very weak. It is therefore unlikely that the attrition in the Netherlands Twin Register affected the generalizability of the results in this thesis.

Response rate in teachers

Next, differences between twins with responding and non-responding teachers are investigated with respect to: i) maternally ratings on AP; ii) maternally ratings on AGG; or iii) maternally ratings on ANX at the same age. These questions were addressed at the ages 7, 10, and 12 years. The results of these analyses are shown in Appendix 2. As with the analyses of the maternal data, the analyses of the problem behavior scores were limited to the data of the first-born twin. Mean maternal AP scores were significantly higher in twins for whom teachers did not return the questionnaire than in twins for whom teachers did return the questionnaire, at the ages 7 and 12 years. Likewise, mean maternal ratings of AGG were higher in twins for whom teacher ratings were available than in twins for whom teacher ratings were not available at age 12 years. No differences were found with respect to the maternal ANX scores. The effect sizes were again very small (.00-.01). Therefore, the fact that we analyzed only the data from the responding teachers, is not likely to bias the results in this thesis.

2. DIAGNOSTIC INTERVIEW DATA COLLECTION

Structured diagnostic interview data were collected via telephone in a sub-sample of the twins of the young NTR to determine the relation between different approaches (e.g., categorical vs quantitative) on the assessment of psychopathology.The diagnostic interview assessed the full spectrum of psychopathology, but my focus was on the assessment and analysis of ADHD. Because only about 5% of the children are expected to have ADHD in the general population, the sample was enriched by selecting children based on their AP scores. The interview data were collected in cooperation with the department of Child and Adolescent Psychiatry at the Erasmus Medical Center/Sophia Children's Hospital in Rotterdam. Mothers of the twins were interviewed by telephone. A total of ten extensively trained medical and psychology students assisted in the collection of the interview data.

Selection criteria

Twins were selected on the basis of their standardized maternal Child Behavior Checklist (CBCL) scores (T-scores; Mean=50; SD=10) at the ages 7, 10, and 12 years. T-scores were computed in boys and girls separately. In other words, girls were selected if they scored low or high compared to other girls, and boys were selected if they scored low or high compared to other boys. This procedure resulted in the selection of an equal number of boys and girls. Subjects were excluded if maternal ratings were available at only a single occasion, or if they suffered from a severe handicap, which disrupts daily functioning (e.g., autism, blindness). Twin-pairs were selected if at least one of the twins scored high on AP (affected pairs), or if both twins scored low on AP (control pairs). A high score was defined as a T-score above 60 at all available time-points (ages 7, 10, and 12 years) and a T-score above 65 at least once. A low score was defined as a T-score below 55 at all available time-points. The control twin-pairs were matched with affected twin-pairs on the basis of sex, birth cohort, maternal age, and Social Economic Status (SES).

Subjects

From 2002 to 2004, 645 mothers of twins born between 1989 and 1994 were invited to participate with the diagnostic interview study. At the time of the interview the twins were between 10 and 13 years old. Of the 645 twin-pairs, 374 were control pairs (both members

scored low), and 271 were affected (at least one of the members scored high). Of the 645 families, 504 (78%) successfully participated (i.e., produced usable data). The response rate was similar in the control pairs (78.5%) and the affected pairs (78.2%). Furthermore, the AP scores in first-born twins showed that the twins whose mothers participated in the interview did not score significantly different on AP than the twins whose mothers did not participate at the ages 7 years ($F(1)=.023$, $p=.879$), 10 years ($F(1)=.038$, $p=.846$), or 12 years ($F(1)=.098$, $p=.754$). This provides support for absence of systematic differences between responders and non-responders with respect to the variable of interest. In Figures 4 and 5, the number of participants are shown by sex and by status of the twin-pair at the time of selection.

Figure 3. Number of twins who participated in the interview study by sex

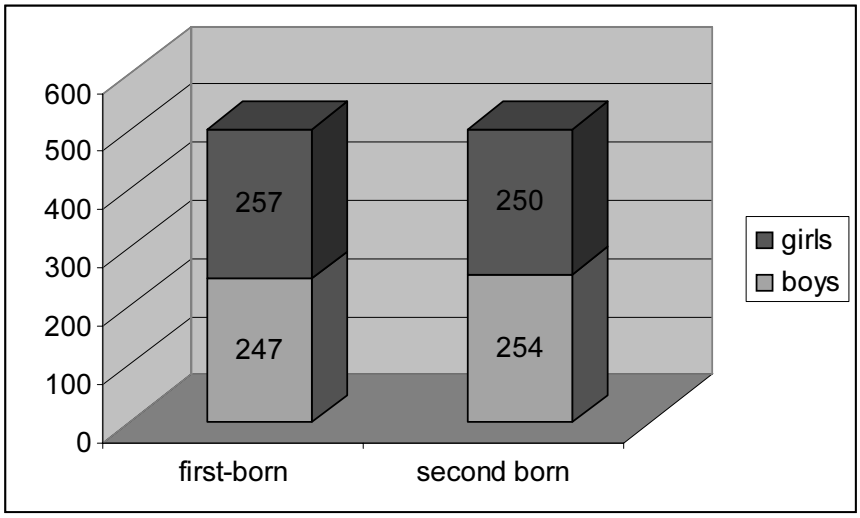
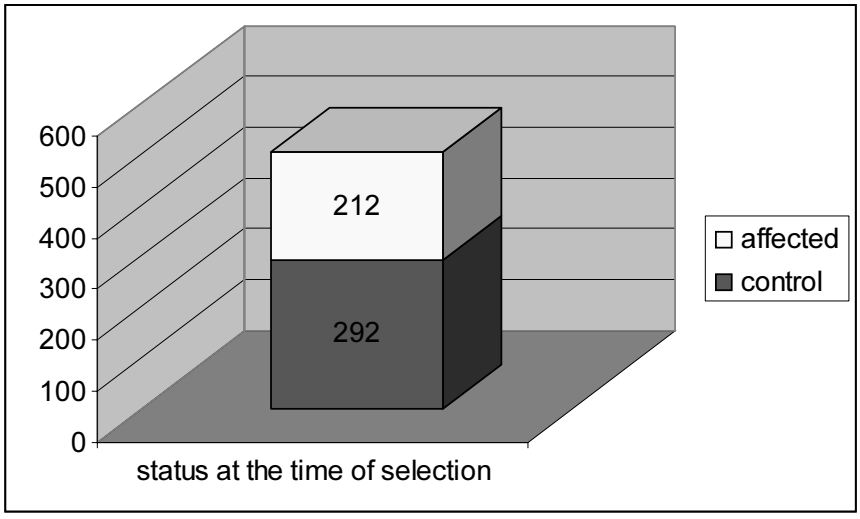


Figure 4. Number of twin-pairs who participated in the interview study by status at the time of selection



Description of interview

The interview consisted of two parts. The first part includes the CBCL-AP items and questions on zygosity, course of pregnancy, medication, sport participation, playing a musical instrument, puberty, education, and social interaction between the two members of the twin-pairs (e.g., whether the twins share a room at home). The second part is the Diagnostic Interview Schedule for Children (DISC) (Shaffer et al., 1993), which is a structured clinical interview based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994). The presence of the following psychiatric diagnoses was assessed: Social phobia, separation anxiety disorder, specific phobia, generalized anxiety disorder, depression, dysthymic disorder, attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder.

Collection of DNA samples

All families that participated in the interview study were asked permission for the collection of DNA from the twins and their parents for zygosity determination and use for future studies (e.g., gene finding studies). Mouth swabs were used to collect buccal cell samples, from which DNA was extracted. Zygosity testing was performed by the Centre for Neurogenomics and Cognition Research (CNCR) at the Vrije Universiteit and included Polymerase Chain Reaction (PCR) of eleven highly polymorphic genetic markers: D12S78, D13S170, D14S1048, D15S153, D19S884, D3S1744, D5S636, D6S1056, D7S821, D8S1133, and D9S1817.

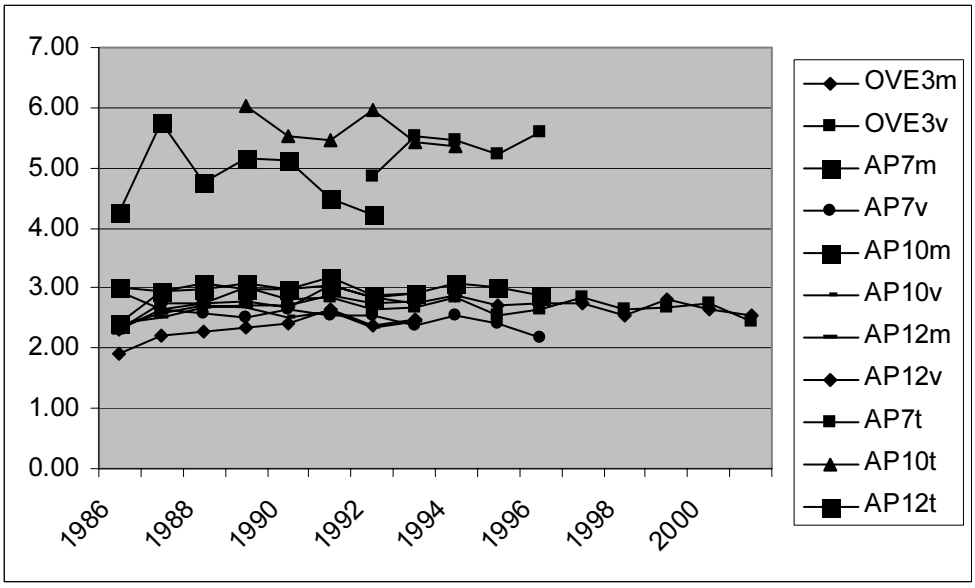
Of the 504 families for whom interview data are available, 368 families granted permission to collect DNA; 115 families did not do so, and 21 families did not respond, or were unable to reach a decision on this matter. Of the 368 families that granted permission to collect DNA, 353 first-borns and 354 second borns provided their DNA samples, and granted permission to determine zygosity. Three hundred and fourty first-borns and 341 second borns also granted permission to use their DNA in future gene finding studies. Of the parents, 314 mothers and 274 fathers provided a DNA-sample, and granted permission to use their DNA in future studies.

At the time of selection, the zygosity of the twins was based on questionnaire data. When DNA became available, the zygosity of 14 of 459 same-sex pairs required revision. Twelve of these pairs appeared to be MZ based on questionnaire data, but were in fact DZ. Two pairs appeared to be DZ based on questionnaire data, but were in fact MZ.

3. THE EFFECT OF BIRTH COHORT ON MEAN LEVELS OF ATTENTION PROBLEMS

It has been suggested that the number of children with high levels of attention problems and hyperactivity has increased in the past decade. If this is true, the data from twins of different birth-cohorts can not be analyzed simultaneously without taking into consideration the mean differences. Therefore, linear regression was used to investigate the effect of birth cohort on mean AP scores in first-born twins. Birth cohort predicted neither the level of parent-rated overactive behavior at age 3, nor parent and teacher rated AP at the ages 7, 10, and 12 years. Figure 3 shows the mean scores as a function of birth cohort. Clearly, combining data from different birth cohorts is justified given the lack of systematic differences between birth cohorts.

Figure 5. The relation between birth cohort and mean Attention Problem (AP) and overactive behavior (OVE) scores as reported by mothers (m), fathers (f), and teachers (t) at the ages 3, 7, 10, and 12 years



GENERAL CONCLUSIONS

In this chapter, I have shown that for the families who have been registered by the NTR, the differences between responding and non-responding parents and teachers are negligible. Also, I did not find an effect of birth cohort on mean levels of AP which justifies collapsing the data from different birth cohorts. These finding support the validity of the results of the previous chapters.

REFERENCE LIST

Achenbach, T. M. (1991a). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991c). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

American Psychiatric Association (1994). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorder: DSM-IV (4th ed.)*. Washington, DC: American Psychiatric Association.

Boomsma, D. I., Vink, J. M., Beijsterveldt van, C. E. M., Geus, E. J. C., Beem, A. L., Mulder, E. J. C. M. et al. (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Centraal Bureau voor de Statistiek (1993). *Standaard Beroepenclassificatie 1992 [Standard Classification of Occupations 1992]*. Heerlen, the Netherlands: Centraal Bureau voor de Statistiek.

Conners, C. K. (2001). *Conners' Rating Scales-Revised*. New York and Toronto: Multi-Health Systems, Inc.

Shaffer, D., Schwabstone, M., Fisher, P., Cohen, P., Piacentini, J., Davies, M. et al. (1993). The Diagnostic Interview Schedule for Children Revised Version (Disc-R) .1. Preparation, Field Testing, Interrater Reliability, and Acceptability. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 643-650.

Spivack, G. & Spotts, J. (1966). *The Devereux Child Behavior (DCB) Rating Scale*. Devon, PA: The Devereux Foundation.

CHAPTER

14

SUMMARY & DISCUSSION

This chapter is based on E.M. Derks, J.J. Hudziak, & Boomsma, D.I. Genetics of ADHD, Hyperactivity, and Attention Problems. *In: Handbook of Behavior Genetics, in press.*

SUMMARY

The aims of this thesis were i) to investigate sex differences in attention problems, hyperactivity and attention deficit hyperactivity disorder (ADHD), ii) to further the development of statistical methods necessary for the analysis of phenotypic data on a variety of measures of ADHD, and iii) to examine the aetiology of individual differences of ADHD and related behavior problems. In this chapter, I first give a summary and discussion of the results that have been described in the previous chapters. I will conclude with some directions for future research.

AIM 1: INVESTIGATION OF SEX DIFFERENCES IN ATTENTION PROBLEMS, HYPERACTIVITY, AND ADHD

Sex differences in ADHD were the main focus of interest in chapters' three to five, with an interest in measurement. Chapter three addressed the question whether the higher prevalence of ADHD and oppositional defiant disorder in boys is the result of different liabilities for these disorders, or, alternatively, is due to measurement bias. To investigate the role of measurement bias, we examined if the four scales of the Conners Teacher Rating Scale (CTRS) (Oppositional behavior, Cognitive problems-Inattention, Hyperactivity, and the ADHD-index) measure the same construct, i.e., the same latent variable of interest, in boys and girls. In addition, we studied the presence of quantitative and categorical sex differences in the genetic and environmental contributions to variation in these behaviors. No sex differences in the factor structure of the CTRS were found, which implies that the CTRS is measurement invariant with respect to sex. The heritability (56-71%) was similar in boys and girls. However, constraining the genetic correlation at .5 in opposite sex twins resulted in a significant decrease in model fit for all four scales which supports the idea that in part different genes contribute to individual differences in problem behavior of boys and girls at school.

In chapter four, we examined the relation between Child Behavior Checklist-attention problem (CBCL-AP) scores and the 4th edition Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) diagnosis of ADHD. Children with a low maternal AP-score obtained a negative ADHD diagnosis in 96% of cases. Children with a high AP-score obtained a positive diagnosis in 36% (girls) and 59% (boys) of cases. The positive predictive power of CBCL-AP was significantly greater in boys than girls. A stronger relationship between CBCL-AP and DSM-IV ADHD in boys than girls was further supported by larger correlations between CBCL-AP and ADHD symptom scores in boys than in girls. In addition, the results of regression analyses show that in children with elevated AP-scores, the number of predicted ADHD symptoms is greater in boys than girls.

Chapter five focuses on the discrepancy of the male:female ratio for ADHD in general populations (3:1) and in clinical settings (9:1). I investigated if this discrepancy was the result of different referral rates in boys and girls. It was shown that boys with ADHD are more often referred for treatment than girls with ADHD. A number of possible explanations for the sex difference in referral rate were investigated. Boys and girls with ADHD showed similar levels of psychiatric illness and school impairment (such as being held back, special class placement, and learning problems) by maternal report. Mothers also reported similar levels of aggression and attention problems in boys and girls with ADHD. In contrast, teachers consistently rated boys with ADHD as having higher scores on reports of attention problems and ADHD related behavior than girls with ADHD. It was concluded that sex differences vary across settings: boys and girls with ADHD are rated as behaving differently at school, but not at home. The higher level of teacher reported problem behavior at school may explain the high male:female ratio for ADHD in clinical settings.

Aim 2: Methodological issues in twin studies on attention problems

A number of methodological issues are particularly relevant to twin studies of behavior problems. The goal of the simulation study in chapter six was to investigate the optimal selection design for the situation in which a cheap measure (X) is observed in a large, representative twin sample and a more expensive measure (Y) is observed in a selected subsample of the twins. This simulation was important to my own study design in which interview data were collected in twins who were selected based on questionnaire data. Scores obtained on X are used to select the most informative twin-pairs with respect to the detection of genetic and environmental influences on the variance of Y. Missingness was introduced in accordance with six selection designs. The percentage of missing data was 88%. It was shown that the selection of informative pairs is most effective when the correlation between traits X and Y is high. The decrease in statistical power as a result of missingness is relatively small. Furthermore, it was found that a design that selects on an individual level rather than on a pair-wise level, is the best design for detecting influences of A, C, and D. However, the extreme discordant and concordant (EDAC) design is to be preferred, if an additional purpose of a study is to detect Quantitative Trait Loci.

Chapter 7 focused on one of the characteristics of rating scales, namely the skewness and kurtosis in the distribution of the sum scores of behavior problems. Genetic and environmental influences on variance in phenotypic traits are usually estimated with normal theory Maximum Likelihood (ML). However, when the assumption of multivariate normality is not met, this method may result in biased parameter estimates and incorrect likelihood ratio tests. When fitting an ACE model to censored data, ignoring the non-normality leads to an unbiased estimate of the additive genetic effects, underestimation of the shared environmental effects, and overestimation of the non-shared environmental effects. Square root transformation and the use of normal scores did not remove this bias. When fitting an ADE model, the additive genetic effect was underestimated, while the dominant and non-shared environmental effects were overestimated. In all models, the correct parameter estimates were recovered with categorical data analysis. However, with categorical data analysis, the statistical power to detect genetic and environmental influences was lower.

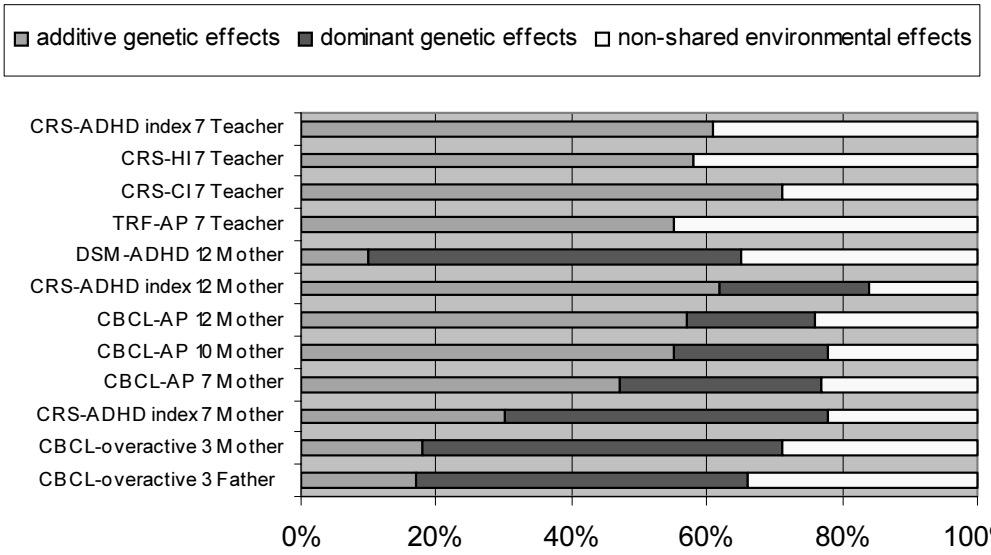
Chapter eight examined a new method to detect violation of the Equal Environment Assumption (EEA) based on multivariate data. In the classic twin design, estimation of genetic and environmental effects is based on the assumption that environmental influences are shared to the same extent by MZ and DZ twins (EEA). We explore the conditions in which the EEA can be tested by estimating the shared environmental correlation in DZ twins using multivariate phenotypic data. Model identification was investigated and the statistical power to detect violation of the EEA was examined in Mx. The amount of bias caused by ignoring violation of the EEA was evaluated. It was shown that bivariate and trivariate models include several instances in which the EEA can be tested. The number of twin-pairs that is needed to detect violation of the EEA with a statistical power of .80 (= .05) varied between 508 and 3576 pairs for the situations considered. The bias in parameter estimates, given misspecification, ranged from 5-34% for additive genetic effects, and from 4-34% for shared environmental effects. Estimates of the non-shared environmental effects were not biased. The analysis of empirical data on spatial ability and aggression indicated no detectable violation of the EEA.

Aim 3: Studying aetiological influences on individual differences in attention problems, hyperactivity, and ADHD

In chapters 3, 9, 10, and 11, we estimated the influences of additive genetic (A), dominant genetic (D), and non-shared environmental influences (E) on individual differences in attention problems, hyperactivity, and ADHD across informants, instrument, and age. Figure 1 summarizes the results of these studies including the results of a study of Hudziak

et al. (2005), in which genetic and environmental influences on maternal Conners Rating Scale-ADHD index scores were estimated using data collected in the Netherlands Twin Registry.

Figure 1. Percentages of total variance explained by additive genetic (A), dominant genetic (D) and non-shared environmental (E) effects in 3, 7, 10, and 12 year old children



Note: CRS=Conners Rating Scale; TRF=Teacher Report Form; DSM=Diagnostic Statistical Manual of Mental Disorders; CBCL=Child Behavior Checklist; AP=attention problems, HI=hyperactivity, ADHD=attention deficit hyperactivity disorder, CI=cognitive problems-inattention

Genetic influences explain the greatest proportion of Individual differences in parental ratings of attention problems (66-84%). Non-shared environmental effects explain the remaining variance. The heritability of teacher ratings on attention problems is somewhat lower, and ranges between 55-71%. Figure 1 shows that the broad-sense heritability does not vary much as a function of age. For example, the broad-sense heritability of the maternal rated CBCL-AP scale is 77% at age 7, 78% at age 10, and 76% at age 12. The amount of total genetic variance attributable to additive (A) versus dominant (D) genetic effects does vary as a function of age and instrument. At age 3, the genetic variance is mainly explained by D, and to a smaller extent by A. As children grow older, the influence of D decreases, and the influence of A increases. An exception is formed by the DSM-ADHD symptom scores, which were collected by means of a structured clinical interview at age 12. Variation in these symptom scores almost entirely results from dominant genetic effects.

In chapter nine, we investigated the degree to which the genetic and environmental effects are moderated by informant in a sample of 9,689 3-year-old twin pairs. Rater Bias and Psychometric Models were fitted to CBCL/2-3 data obtained from mothers and fathers to determine the genetic and environmental contributions to five CBCL syndromes: aggressive, oppositional, overactive, withdrawn, and anxious/depressed behavior. Parental ratings were found to be influenced by aspects of the child's behavior that are experienced in the same way by both parents and by aspects of the child's behavior that are experienced uniquely by each parent. There is evidence for large genetic contributions to individual dif-

ferences in all CBCL syndromes. Shared and non-shared environmental influences play significant roles as well. One exception is overactive behavior, which is only influenced by genetic and non-shared environmental influences.

In chapter 10, the genetic and environmental influences on maternal and teacher ratings of Attention Problems (AP) were examined in 7-year-old children. Teachers completed the Teacher Report Form (N=2259 pairs), and mothers completed the Child Behavior Checklist (N=2057 pairs). Higher correlations were found in twins rated by the same teacher than in twins rated by different teachers. This was explained by a correlated error in same teacher ratings. This correlated error is either the result of i) raters having their own subjective view on which behaviors are appropriate and which are not, or ii) the influence that a rater has on the behavior of the child because of the rater's own personality characteristics. We further found that 41% of the variance in maternal and teacher ratings is explained by a common factor. The broad-sense heritability of this common factor is 78%. The broad-sense heritabilities of the rater specific factors of mothers and teachers are 76% and 39%, respectively. The fact that the rater specific factors are heritable indicates that mothers and teachers assess unique aspects of the child's behavior.

The aim of chapter 11 was to investigate the extent to which individual differences in maternal ratings on three different instruments that assess attention problems and ADHD, reflect the same genetic and environmental influences. The total sample for whom at least one maternal ratings was available consisted of 10916 twins from 5458 families. Child Behavior Checklist (CBCL) ratings were available for 10018, 6565, and 5780 twins at the ages 7, 10, and 12, respectively. The Conners Rating Scale (4887 twins) and the DSM interview (1006 twins) were completed at age 12. Statistical analyses showed phenotypic correlations of the three instruments in the range of .45 to .77. The variances and covariances of the five instruments were mainly explained by genetic influences. The genetic correlations of the data collected at age 12 varied between .61 and 1.00. In conclusion, the genetic overlap between behavior checklist scores and the DSM-IV diagnosis of ADHD is large. This implies that the costs of gene finding studies can be reduced by collecting survey data instead of diagnostic interview data.

The objective of chapter 12 was to investigate the specific environmental influences that play a role for attention deficit hyperactivity disorder (ADHD). A sample of twins was followed longitudinally from birth onwards. Questionnaires were completed by mothers (ages 1, 2, 3, 7, 10, and 12 years), teachers (age 7, 10, and 12 years), and the twins themselves (12 years). Mothers completed a structured clinical interview. From the large longitudinal sample, discordant monozygotic twin-pairs were selected and matched with concordant pairs. The final sample (95 pairs) consists of discordant, concordant-high, and concordant-low twin-pairs. Child-specific environmental influences were compared between the affected and unaffected twins within the discordant pairs. Environmental factors that are shared by the two members of a twin-pair (e.g., maternal smoking during pregnancy) were compared between the discordant, concordant-high and concordant-low groups. The affected members of the discordant twin-pairs scored higher than the unaffected members on a wide variety of measures of AP, ADHD, and other behavior problems according to the mother, teacher and self. Affected members had lower birth weight and poorer early motor development than unaffected members. Differences between discordant, concordant-low, and concordant-high groups were reported for maternal smoking, sleeping in different rooms, and living with only one of the parents. Maternal smoking was more common in the concordant-high group than in the concordant-low group. The discordant twins did not differ from the concordant twins. Sleeping in separate rooms was more common in discordant pairs than in concordant low pairs at the ages 2-6 years. More concordant-high pairs lived with only one parent than concordant-low or discordant pairs.

DISCUSSION

AIM 1: INVESTIGATION OF SEX DIFFERENCES IN ATTENTION PROBLEMS, HYPERACTIVITY, AND ADHD

The four Conners Teacher Rating Scales, which measure ADHD and related behavior problems, are not biased with respect to sex in 7 year-olds. Future studies will reveal if this finding can be generalized to parental ratings, to children of different age groups, and to different rating scales. In studies of groups differences in genetic covariance structure, measurement invariance (i.e., unbiased measurement) is an important issue, because any established group difference cannot be interpreted unambiguously if measurement invariance does not hold (Lubke et al., 2004). Differences between boys and girls were reported in a number of areas. First, variation in teacher, but not parental ratings, is influenced by partly different genes in boys and girls. Second, the relation between Child Behavior Checklist-AP scores and DSM-IV ADHD is stronger in boys than girls. Third, boys with ADHD have a higher chance of being referred for treatment than girls with ADHD, which is probably caused by higher levels of problem behavior at school in boys than girls.

Sex differences in behavior vary as a function of the context in which children are rated. Parents report similar mean levels of attention problems (AP) and aggression (AGG) in boys and girls with ADHD. In contrast, teachers report higher levels of AP and AGG in boys than in girls. The lower level of problem behavior at school in girls with ADHD suggests that girls adapt better to the school environment. Why would this be the case? Abikoff et al. (2002) observed the classroom behavior of children with ADHD. They showed that boys with ADHD had higher rates of interference, total aggression, and gross motor movements than girls with ADHD. Girls and boys did not differ in their rates of off-task and minor motor movements. The expression of ADHD seems to vary as a function of the child's sex. While boys and girls are just as likely to be inattentive, boys show more disruptive behavior. In agreement with these findings, in chapter three we reported similar teacher rated levels of cognitive problems-inattention in boys and girls, while boys scored higher on oppositional behavior, hyperactivity, and the ADHD-index. This shows that the sex differences may be present not only in children with ADHD, but also in the general population.

The lower level of disruptive behavior in girls may explain their lower referral rate compared to boys. Although the fact that a girl is less disruptive in school may in some instances be a valid reason for not referring her for treatment, this may not always be true. In chapter three, it was reported that boys and girls with ADHD show similar levels of cognitive problems/inattention. Furthermore, Dalsgaard et al. (2002) showed that girls with ADHD more often suffered from psychiatric problems as adults than boys with ADHD. Finally, Berry et al. (1985) reported that among children with ADHD, girls were more likely to suffer peer rejection than boys. Therefore, in deciding whether to refer a girl with a high level of attention problems for treatment, it should be acknowledged that she may not show a high level of disruptive behavior, but that she might instead have learning problems, social problems, or experience peer rejection. An additional concern in clinical decision making is that the DSM diagnosis of ADHD may be biased with respect to sex. The fact that the association between the CBCL and DSM is stronger in boys than girls could imply that the DSM includes symptoms that are more relevant for boys than girls. If this is the case, the DSM symptoms would not be measurement invariant with respect to sex. One of our future goals is to study measurement invariance with respect to sex for DSM-ADHD.

Another context-dependent sex difference is apparent in the evaluation of qualitative sex differences in parental and teacher ratings on attention problems. In parent ratings, the same genes contribute to variance in boys and girls. In contrast, the factors that contribute to the variance in teacher ratings vary as a function of sex. These factors appear to be genetic rather than environmental.

Aim 2: Methodological issues in twin studies on attention problems

ADHD and the Equal Environment Assumption

In chapter eight, we explored a test of the Equal Environment Assumption (EEA) in multivariate data, and we applied this method to measures of spatial ability and aggression. Because of the focus of this thesis, it may have been surprising that we did not apply this method to ADHD. The reason that we chose a different example is that ADHD is influenced by dominant genetic influences. In the classical twin design, shared environmental and dominant genetic effects can not be simultaneously estimated. Although the DZ correlations, which are less than half the MZ correlations may also be explained by violation of the EEA instead of dominance, the shared environment should correlate less than .5 in DZ twins to explain this finding. Because this is quite unlikely intuitively, ADHD is not a very good example for illustrative purposes. The present exercise in testing the EEA in the classical twin design exploits the presence of multivariate data. It would seem that multivariate data provide greater possibilities to test such assumptions than have to date been realized. This remains an interesting avenue for future study.

Genetic model fitting analyses: interpretation of the factors

One of the limitations in chapter eight concerned the interpretation of violation of the Equal Environment Assumption (EEA). A shared environmental correlation smaller than 1 in DZ twins could be interpreted as a violation of the EEA, but could also be the result of a different process (e.g., assortative mating). The difficulty arises because the interpretation is based on the correlational structure of the latent factors. This problem is not unique to behavioral genetic studies. In exploratory factor analytic studies, variation in item scores is explained by one or more latent factors. The name of a factor is chosen based on interpretation of the specific items that load highly on this factor (Stevens, 1996). In behavioral genetic studies, the factors are defined based on theoretical assumptions about the correlational structure of genetic and environmental influences. For example, provided the absence of assortative mating, segregating genes correlate 1 in MZ twins and .5 in DZ twins. However, there may be a different factor that also correlates 1 in MZ twins and .5 in DZ twins. If this is the case, these two factors can not be distinguished from each other, because of the identical expectation of the twin covariances. Does this possible problem of interpretation detract from the usefulness of behavioral genetic studies? I do not believe that this is the case.

A couple of years ago, Joseph (2000) criticized the results of twin studies on ADHD. He stated that the validity of the EEA assumption has not been proven, and that environmental influences may account for the greater phenotypic similarity in MZ than DZ twins. A challenging reply was given by Faraone & Biederman (2000) who argued that, although Joseph rejects the standard interpretation of the results of twin studies, he fails to provide an alternative theory with testable predictions. Such a testable theory would need to specify an environmental factor that i) is transmitted from parents to children; ii) is more likely to be shared by MZ than DZ twins; iii) explains the elevated rates of ADHD and associated traits among the biological relatives of adopted away ADHD children. A genetic theory of ADHD explains these findings with a single idea, while Joseph appeals to a number of different mechanisms to explain these findings, and does not describe nor tests the predictions based on the proposed mechanisms. Therefore, there is no reason to reject the current theory of genetic influences on individual differences in ADHD, although we should be willing to consider and test alternative theories that include testable predictions.

AIM 3: STUDYING AETIOLOGICAL INFLUENCES ON INDIVIDUAL DIFFERENCES IN ATTENTION PROBLEMS, HYPERACTIVITY, AND ADHD

Genetic and environmental effects

Irrespective of informant, age of the child, and measurement instrument, the main conclusion of this thesis is that ADHD is largely influenced by genetic influences, and non-shared environmental influences. Shared environmental influences do not seem to be present, although it is possible that gene-environment interaction plays a role. Variation due to interaction between genetic factors and shared environmental factors can not be distinguished from genetic variance in the classical twin design. Gene-environment interaction has been found for maternal smoking (Kahn et al., 2003), which suggests that the higher prevalence of maternal smoking in the concordant-high group than in the concordant-low group reported in chapter 12 also has a direct influence on ADHD. I elaborate on gene-environment interaction in the directions for future research section. As modelled in the present thesis, the genetic influences comprise additive genetic (A) and dominant genetic (D) effects. The finding of dominant genetic influences is inconsistent with the findings of Rietveld et al. (2003), who reported the influence of social interaction. The fact that no variance differences were observed in MZ and DZ twins in a sample that was more than twice as large as the Rietveld et al. sample, suggests that at age 3, genetic dominance is the more plausible hypothesis. Does the high heritability of ADHD have any implications for clinical interventions? Yes and no. It may have future implications when more is known about the specific genes that underlie individual differences in AHDH, and about the biological pathways involved. For example, an understanding of the latter may give rise to new medications. It should be emphasized, however, that a high heritability does not imply that environmental interventions cannot be effective. A well known example of a heritable trait of which the outcome is amenable to environmental intervention is phenylketonuria (pku), an autosomal recessive trait. Although the predisposition for this disease is 100% heritable, the adverse consequences of pku can be avoided by an environmental intervention in the form of a diet. Therefore, the fact that ADHD is a heritable trait should not discourage researchers and clinicians from considering environmental interventions.

Non-shared environmental factors also contribute to individual differences in attention problems. In chapter 12, we used the discordant MZ twin design to identify several environmental factors that are associated with attention problems. Discordant MZ pairs were found to differ in birth weight and early motor development, which may prove to be significant markers of ADHD. Because differences in MZ twins can not result from genetic differences, this shows that the differences are truly environmental in nature. As the knowledge on the relevant genetic and environmental factors increases, future studies on ADHD may focus on the complex interplay between genetic and environmental factors. We will elaborate on this below.

Multi-informant models

Paternal, maternal, and teacher ratings of AP show moderate to high correlations. We estimated the extent to which the agreement among different raters is caused by the same genetic and/or environmental factors. The factor that was common to maternal and paternal ratings on overactivity explained 68% of the variance, and the factor that was common to maternal and teacher ratings on attention problems explained 41% of the variance. This confirms the findings of Achenbach & Rescorla (2001), who reported a higher correlation of maternal and paternal ratings than of parental and teacher ratings. The variance of the common factors was mainly explained by genetic dominance effects (D), rather than additive genetic effects (A).

The multi-informant studies show that individual raters offer unique perspectives on the presence of attention problems in children. Not only is it true that children's behavior depends on the situation, their behavior may also vary as a function of the person

with whom they interact. It is difficult to disentangle these two processes. Do fathers and mothers rate different aspects of their children's behavior, because they interact differently with their children, possibly as a function of their own personality characteristics, or do they rate different aspects, because they observe them in different situations? A similar question arises in the interpretation of the higher correlations in twins rated by the same teacher than in twins rated by different teachers. Is this the result of teacher rater bias, and are the two groups just as alike, or are twins who are placed in the same classroom actually more alike as a result of greater environmental similarity? The fact that the maternal correlations are not higher in twins, who are placed into the same classroom, compared to twins who are placed in different classrooms, shows that, if the latter explanation is true, the environmental effect is limited to the school environment. With the current data we can not distinguish between these alternative explanations. One possibility to investigate this question is by observation studies of twins in the same and in different classrooms.

DIRECTIONS FOR FUTURE RESEARCH

In addressing the three main research questions in this thesis, a number of new issues emerged. In this section we will identify the questions that have yet to be answered, and provide some suggestions for future research.

Phenotype definitions

In many instances, the heritability of a trait is based on the analysis of sum scores (e.g., of items or symptoms), and the distribution of the sum scores often displays a large degree of skewness and kurtosis. Especially when analyzing symptom data on psychopathology, the distribution of sum scores is often L-shaped, due to the fact that the vast majority of subjects displays few or no symptoms (Oord van den et al., 2003). In a simulation study of such data, I showed that if the true model is an ADE model, and parameters are estimated with normal theory Maximum Likelihood, the additive genetic component is underestimated, and the non-additive genetic component and the non-shared environmental component are overestimated (Derks et al., 2004). The use of a liability threshold model (Lynch & Walsh, 1998) when analyzing sum scores with an L-shaped distribution was recommended.

Another concern in analyzing sum scores that is not resolved by using a threshold model for the sum scores, is that some of the information that is contained in the original item scores is lost in the analysis of sum scores (Neale et al., 2005). The fact that the relationship between the latent trait and the observed item score may well be probabilistic (i.e., a person, who is below the threshold on the latent trait, has a relatively low probability to score positive on the item), instead of deterministic (i.e., a person who is below the threshold, has a zero probability to score positive on the item) may also cause bias in the heritability estimate. Within the Item Response Theory (IRT) framework, item scores are modeled as a function of one or more latent factors. Two recent papers show the advantages of IRT in the behavior genetic research field (Eaves et al., 2005; Berg van den et al., in press). According to Berg van den et al., advantages of using an IRT framework include the fact that IRT provides an explicit model for the relation between item scores and the latent phenotype, and it supports the use of incomplete item administration and handling of missing data. The application of this approach in future studies on ADHD is particularly interesting for gene finding studies while it may increase statistical power to detect the influence of genes with small effects.

Heritability of ADHD in adults

The heritability of ADHD has been studied extensively in children. In contrast, little is known about the magnitude of the genetic and environmental influences on individual differences in ADHD in adults. This may partly be explained by the fact that some of the earlier work suggested that ADHD is rare in adulthood. However, Faraone et al. (2005) per-

formed a meta-analysis of follow-up studies on ADHD. They show that syndromatic persistence (i.e., the maintenance of full diagnostic status) is low (~15%), but that symptomatic persistence (i.e., the maintenance of partial diagnostic status with impairment) is much higher with a persistence rate of 40-60%. Therefore, future research should focus on the identification of the genetic and environmental influences on individual differences in ADHD in adults. The only study of the aetiology of individual differences in attention problems in adults was based on self report data from the Netherlands Twin Register at three different time waves (Berg van den et al., 2006). The mean age of the young adults is 19.6, 21.3, and 22.8 years at wave 1, 2, and 3, respectively. Irrespective of measurement wave, the heritability of attention problems is about 40%. Berg van den et al. (2006) further showed that the stability in attention problems is largely due to genetic factors. In addition, variance in ADHD at different ages in young adulthood, is due to the effects of the same genes. It is unclear if the lower heritabilities in adults, compared to children, can be explained by age effects, or by the fact that ratings of ADHD are usually based on parental or teacher reports in children and on self-reports in adults. Future studies of the Netherlands Twin Register will look into genetic and environmental influences on stability of the attention problems from early childhood (parent and teacher reports), through adolescence (parent, teacher, and self-reports) into adulthood (self-reports).

Genotype-environment correlation and genotype-environment interaction

Genetic factors contribute to individual differences in organisms that are characterized by and subject to a particular range of environmental effects. Likewise, environmental effects operate on organisms that differ genetically. In giving rise to phenotypic individual differences, genetic and environmental effects do not necessarily operate independently. Notably, mutual dependency may result in genotype-environment correlations. Genotype-environmental correlations refer to situations in which individual differences in genetic liability are associated with individual differences in specific environmental circumstances. Genotype-environmental interactions refer to a type of genetic liability that involves a differential susceptibility to specific environmental features (Rutter et al., 1997).

As far as we know, genotype-environment correlations have not been studied for ADHD. Genotype-environment interactions have received more attention. Initially, no evidence was found for genotype-environment interaction in a study of interaction effects between parental ADHD status (i.e., a marker for genetic risk) and a large number of environmental risk factors (Milberger et al., 1997). Recent findings, however, support the presence of environmental mediation of the effect of two candidate genes for ADHD. Kahn et al. (2003) investigated GxE interaction between the DAT genotype and prenatal smoke exposure. They showed that children with both prenatal smoke exposure and the DAT +/+ genotype had significantly elevated hyperactive/impulsive scores compared with children with no smoke exposure and DAT +/- or DAT -/- genotypes. Inattentive scores were not significantly elevated. No main effects of either prenatal smoke exposure or DAT +/+ genotype were found. An interaction between the DRD4-7-repeat (DRD4*7) allele and the season of birth was reported for hyperkinetic disorder + conduct disorder (HD+CD) (Seeger et al., 2004). The presence of one copy of the DRD4*7 decreases the risk of HD+CD in children born in autumn and winter while it increases the risk of HD+CD in children born in spring and summer. One hypothetical consideration is that autumn and winter births correlate with longer day lengths and decreased melatonin secretion. There might be an interaction between the day length, the melatonin-dopamine system and the subsensitive postsynaptic receptor of the DRD4*7 allele. The finding of significant GxE interaction in these studies highlights the importance of considering the effects of both environmental and genetic factors, and their interactions in future studies on ADHD.

It was recently shown how environmental moderator variables can be included in genetic model fitting analyses in a manner that allows for testing of nonlinear moderator properties, and genotype-environment interaction in the presence of genotype-environment correlation (Purcell, 2002). In the NTR, we are currently collecting data on environmental indicators that may moderate the genetic effects on ADHD. In the future, we will study genotype-environment correlations and interactions.

Psychiatric comorbidity in children with ADHD

In this thesis, I have largely ignored the fact that children with ADHD may show higher levels of other psychiatric disorders than controls. A number of twin studies have examined the genetic and environmental effects on the covariation between ADHD and oppositional defiant disorder (ODD) and/or conduct disorder (CD). It was found that variance in ADHD and CD is mainly explained by genetic factors while variance in ODD was explained by both genetic and shared environmental factors (Burt et al., 2001). The covariance of ADHD, CD, and ODD was mainly accounted for by shared environmental factors (47-59%), and also by genetic factors (22-35%). These findings were contradicted by the study of Nadder et al. (2002) who showed that the covariance of ADHD and ODD/CD was only explained by genetic factors and not by environmental factors. A third twin study on the relation between ADHD, ODD, and CD was performed by Dick et al. (2005). The main difference with the previous studies is the focus on self-report data instead of parental or teacher data. This may lead to different results because the correlations between parental and self-report data are usually low. However, the self-report data show the same pattern as the study of Nadder et al.; the genetic overlap between the three disorders is high. The genetic correlations are .46 (ADHD & CD), .74 (ADHD & ODD), and .58 (CD & ODD). While the sample sizes of the previous studies are relatively small (~700 pairs), a questionnaire based study in 2082 twin-pairs showed that the overlap of ADHD and CD was explained by genetic and non-shared environmental factors (Thapar et al., 2001). In summary, there is overwhelming evidence that the association between ADHD, ODD, and CD is mainly due to overlap in genetic factors.

In a recent review of predictors of antisocial behavior in children with ADHD, Thapar et al. (2006) found that the genetic influences on antisocial behavior are the same as those that influence ADHD although these authors emphasized the importance of further study of the interaction between genes and environment. In contrast to the efforts spent to disentangle the relation between ADHD, ODD, and CD, the aetiology of the covariance of ADHD and internalizing disorders, such as depression, has received much less attention. Willcutt et al. (1999) showed that symptoms of inattention were associated with higher levels of depression, while symptoms of hyperactivity were associated more strongly with ODD and CD. Therefore, it is worthwhile to study the association between ADHD and other behavior problems by subtype in a genetically informative design. One of my aims is to reveal the genetic and environmental influences on the covariance of ADHD and depression, and also to perform trivariate analyses of ADHD, disruptive behaviors (i.e., aggression, ODD, or CD), and internalizing disorders.

REFERENCE LIST

Abikoff, H. B., Jensen, P. S., Arnold, L. L. E., Hoza, B., Hechtman, L., Pollack, S. et al. (2002). Observed classroom behavior of children with ADHD: Relationship to gender and comorbidity. *Journal of Abnormal Child Psychology, 30,* 349-359.

Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Department of Psychiatry.

Berg van den, S. M., Glas, C. A. W., & Boomsma, D. I. (in press). Variance decomposition using an IRT measurement model. *Behavior Genetics.*

Berg van den, S. M., Willemsen, G., de Geus, E. J., & Boomsma, D. I. (2006). Genetic etiology of stability of attention problems in young adulthood. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 141,* 55-60.

Berry, C. A., Shaywitz, S. E., & Shaywitz, B. A. (1985). Girls with Attention Deficit Disorder - A Silent Minority - A Report on Behavioral and Cognitive Characteristics. *Pediatrics, 76,* 801-809.

Burt, S. A., Krueger, R. F., McGue, M., & Iacono, W. G. (2001). Sources of covariation among attention-deficit/hyperactivity disorder, oppositional defiant disorder, and conduct disorder: The importance of shared environment. *Journal of Abnormal Psychology, 110,* 516-525.

Dalsgaard, S., Mortensen, P. B., Frydenberg, M., & Thomsen, P. H. (2002). Conduct problems, gender and adult psychiatric outcome of children with attention-deficit hyperactivity disorder. *British Journal of Psychiatry, 181,* 416-421.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research, 7,* 659-669.

Dick, D. M., Viken, R. J., Kaprio, J., Pulkkinen, L., & Rose, R. J. (2005). Understanding the covariation among childhood externalizing symptoms: Genetic and environmental influences on conduct disorder, attention deficit hyperactivity disorder, and oppositional defiant disorder symptoms. *Journal of Abnormal Child Psychology, 33,* 219-229.

Eaves, L., Erkanli, A., Silberg, J., Angold, A., Maes, H. H., & Foley, D. (2005). Application of Bayesian inference using Gibbs sampling to item-response theory modeling of multi-symptom genetic data. *Behavior Genetics, 35,* 765-780.

Faraone, S. V. & Biederman, J. (2000). Nature, nurture, and attention deficit hyperactivity disorder. *Developmental Review, 20,* 568-581.

Faraone, S. V., Biederman, J., & Mick, E. (2005). The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological Medicine, 35,* 1-7.

Hudziak, J. J., Derks, E. M., Althoff, R. R., Rettew, D. C., & Boomsma, D. I. (2005). The genetic and environmental contributions to attention deficit hyperactivity disorder as measured by the Conners' Rating Scales - Revised. *American Journal of Psychiatry, 162,* 1614-1620.

Joseph, J. (2000). Not in their genes: A critical view of the genetics of attention-deficit hyperactivity disorder. *Developmental Review, 20,* 539-567.

Kahn, R. S., Khoury, J., Nichols, W. C., & Lanphear, B. P. (2003). Role of dopamine transporter genotype and maternal prenatal smoking in childhood hyperactive-impulsive, inattentive, and oppositional behaviors. *Journal of Pediatrics, 143,* 104-110.

Lubke, G. H., Dolan, C. V., & Neale, M. C. (2004). Implications of absence of measurement invariance for detecting sex limitation and genotype by environment interaction. *Twin Research, 7,* 292-298.

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

Milberger, S., Biederman, J., Faraone, S. V., Guite, J., & Tsuang, M. T. (1997). Pregnancy, delivery and infancy complications and attention deficit hyperactivity disorder: Issues of gene-environment interaction. *Biological Psychiatry, 41,* 65-75.

Nadder, T. S., Rutter, M., Silberg, J. L., Maes, H. H., & Eaves, L. J. (2002). Genetic effects on the variation and covariation of attention deficit-hyperactivity disorder (ADHD) and oppositional-defiant disorder/ conduct disorder (ODD/CD) symptomatologies across informant and occasion of measurement. (vol 32, pg 39, 2002). *Psychological Medicine, 32,* 378.

Neale, M. C., Lubke, G., Aggen, S. H., & Dolan, C. V. (2005). Problems with using sum scores for estimating variance components: Contamination and measurement noninvariance. *Twin Research and Human Genetics, 8,* 553-568.

Oord van den, E. J. C. G., Pickles, A., & Waldman, I. (2003). Normal variation and abnormality: an empirical study of the liability distributions underlying depression and delinquency. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 44,* 180-192.

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5,* 554-571.

Rietveld, M. J. H., Hudziak, J. J., Bartels, M., Beijsterveldt van, C. E. M., & Boomsma, D. I. (2003). Heritability of attention problems in children: I. Cross-sectional results from a study of twins, age 3-12 years. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 117B,* 102-113.

Rutter, M., Dunn, J., Plomin, R., Simonoff, E., Pickles, A., Maughan, B. et al. (1997). Integrating nature and nurture: Implications of person-environment correlations and interactions for developmental psychopathology. *Development and Psychopathology, 9,* 335-364.

Seeger, G., Schloss, P., Schmidt, M. H., Ruter-Jungfleisch, A., & Henn, F. A. (2004). Gene-environment interaction in hyperkinetic conduct disorder (HD+CD) as indicated by season of birth variations in dopamine receptor (DRD4) gene polymorphism. *Neuroscience Letters, 366,* 282-286.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. (3 ed.) Mahwah: Lawrence Erlbaum Associates, Inc.

Thapar, A., Harrington, R., & McGuffin, P. (2001). Examining the comorbidity of ADHD-related behaviours and conduct problems using a twin study design. *British Journal of Psychiatry, 179,* 224-229.

Thapar, A., Bree van den, M., Fowler, T., Langley, K., & Whittinger, N. (2006). Predictors of antisocial behaviour in children with attention deficit hyperactivity disorder. *European Child & Adolescent Psychiatry, 15,* 118-125.

Willcutt, E. G., Pennington, B. F., Chhabildas, N. A., Friedman, M. C., & Alexander, J. (1999). Psychiatric comorbidity associated with DSM-IV ADHD in a nonreferred sample of twins. *Journal of the American Academy of Child and Adolescent Psychiatry, 38,* 1355-1362.

APPENDICES

APPENDIX 1

PARENTAL RESPONSE RATE, PARENTAL PROBLEM BEHAVIOR SCORES OF THE FIRST-BORN TWINS, AND LEVEL OF SES AS A FUNCTION OF THE PARENTAL (NON)RESPONSE AT THE PREVIOUS AGE

DEPENDENT VARIABLE	INDEPENDENT VARIABLE	LEVELS OF INDEPENDENT VARIABLE	N	SUMMARY STATISTIC	TEST STATISTIC	PARTIAL η^2	P-VALUE
Age 3; response	Age 2	Response	17747	78%	$\chi^2(1)$ =4463	-	<.001
		Non-response	4502	26%			
Age 3; OVE	Age 2	Response	11559	2.76 (2.18)	F(1)=1.01	.00	.315
		Non-response	943	2.84 (2.30)			
Age 3; AGG	Age 2	Response	11543	3.35 (2.79)	F(1)=2.17	.00	.141
		Non-response	937	3.58 (2.98)			
Age 3; ANX	Age 2	Response	11484	3.64 (3.09)	F(1)=.02	.00	.886
		Non-response	931	3.56 (3.14)			
Age 3; SES	Age 2	Response	9077	3.15 (.89)	F(1)=3.50	.00	.062
		Non-response	747	3.22 (.93)			
Age 5; response	Age 3	Response	13985	76%	$\chi^2(1)$=4429	-	<.001
		Non-response	5467	24%			
Age 5; AP	Age 3	Response	8297	11.58 (3.60)	F(1)=2.03	.00	.154
		Non-response	940	11.75 (3.72)			
Age 5; AGG	Age 3	Response	8311	12.20 (3.68)	F(1)=1.11	.00	.292
		Non-response	944	12.35 (3.84)			
Age 5; ANX	Age 3	Response	8310	10.88 (3.37)	F(1)=1.44	.00	.230
		Non-response	946	11.04 (3.58)			
Age 7; response	Age 5	Response	10110	79%	$\chi^2(1)$=4086	-	<.001
		Non-response	5026	25%			
Age 7; AP	Age 5	Response	6513	2.93 (2.90)	F(1)=11.89	.00	.001
		Non-response	963	3.28 (3.19)			
Age 7; AGG	Age 5	Response	6500	6.79 (5.84)	F(1)=9.19	.00	.002
		Non-response	960	7.40 (6.24)			
Age 7; ANX	Age 5	Response	6501	2.31 (2.83)	F(1)=14.24	.00	<.001
		Non-response	960	2.68 (3.20)			
Age 7; SES	Age 5	Response	6535	3.35 (.88)	F(1)=18.14	.00	<.001
		Non-response	987	3.21 (.91)			
Age 10; response	Age 7	Response	7066	77%	$\chi^2(1)$=2853	-	<.001
		Non-response	3408	22%			
Age 10; AP	Age 7	Response	3977	2.92 (3.03)	F(1)=12.75	.00	<.001
		Non-response	536	3.45 (3.24)			
Age 10; AGG	Age 7	Response	3961	6.12 (5.83)	F(1)=5.80	.00	.016
		Non-response	534	6.77 (6.21)			
Age 10; ANX	Age 7	Response	3973	2.67 (3.36)	F(1)=1.25	.00	.264

Note:OVE=overactive; AP=Attention Problems; AGG=Aggression; ANX=Anxiety; SES=Social Economic Status
Partial η^2=effect size (proportion of total variability attributable to the factor)

(Continued)

Dependent variable	Independent variable	Levels of independent variable	N	Summary statistic	Test statistic	Partial η^2	p-value
		Non-response	534	2.84 (3.47)			
Age 10; SES	Age 7	Response	4005	3.32 (.92)	F(1)=9.09	.00	.003
		Non-response	548	3.20 (.93)			
Age 12; response	Age 10	Response	4772	78%	$\chi^2(1)=2351$	-	<.001
		Non-response	2404	18%			
Age 12; AP	Age 10	Response	3627	2.67 (2.87)	F(1)=4.80	.00	.029
		Non-response	409	3.03 (3.20)			
Age 12; AGG	Age 10	Response	3621	5.35 (5.43)	F(1)=2.76	.00	.097
		Non-response	405	5.83 (5.58)			
Age 12; ANX	Age 10	Response	3625	2.43 (3.25)	F(1)=.448	.00	.503
		Non-response	408	2.57 (3.33)			

Note:OVE=overactive; AP=Attention Problems; AGG=Aggression; ANX=Anxiety; SES=Social Economic Status
Partial η^2=effect size (proportion of total variability attributable to the factor)

APPENDIX 2

Parental problem behavior scores of the first-born twins as a function of the (non)response of the teacher at the same age

Dependent variable	Independent variable	Levels of independent variable	N	Summary statistic (% or mean (sd))	Test statistic	Partial η^2	p
Age 7; AP parent	Age 7	Responding teacher	2256	2.78 (2.81)	F(1)=16.09	.01	<.001
		Non-responding teacher	552	3.34 (3.13)			
Age 7; AGG parent	Age 7	Responding teacher	2253	6.63 (5.83)	F(1)=2.48	.00	.115
		Non-responding teacher	550	7.07 (5.97)			
Age 7; ANX parent	Age 7	Responding teacher	2254	2.25 (2.78)	F(1)=3.64	.00	.056
		Non-responding teacher	552	2.51 (3.02)			
Age 10; AP parent	Age 10	Responding teacher	1827	2.93 (2.97)	F(1)=2.34	.00	.126
		Non-responding teacher	513	3.12 (3.20)			
Age 10; AGG parent	Age 10	Responding teacher	1822	6.23 (6.00)	F(1)=2.38	.00	.123
		Non-responding teacher	509	6.69 (5.98)			
Age 10; ANX parent	Age 10	Responding teacher	1824	2.67 (3.39)	F(1)=1.50	.00	.221
		Non-responding teacher	513	2.85 (3.48)			
Age 12; AP parent	Age 12	Responding teacher	1870	2.65 (2.86)	F(1)=7.79	.00	.005
		Non-responding teacher	733	3.00 (3.02)			
Age 12; AGG parent	Age 12	Responding teacher	1867	5.29 (5.31)	F(1)=8.18	.00	.004
		Non-responding teacher	733	5.97 (5.69)			
Age 12; ANX parent	Age 12	Responding teacher	1870	2.45 (3.22)	F(1)=1.37	.00	.243
		Non-responding teacher	733	2.62 (3.30)			

Note:AP=Attention Problems; AGG=Aggression; ANX=Anxiety

SAMENVATTING

SAMENVATTING

Dit proefschrift, met als titel: "Meetproblemen en de genetische invloed op concentratie-problemen, hyperactiviteit en aanverwante stoornissen" bestaat uit drie delen. Deze drie delen corresponderen met drie verwante onderzoeksdoelen: 1) Het onderzoeken van sekseverschillen in concentratie-problemen, hyperactiviteit en attention deficit hyperactivity disorder (ADHD); 2) Het ontwikkelen van statistische methoden voor de analyse van de verzamelde gegevens met betrekking tot ADHD; 3) Het onderzoeken van genetische en omgevingsinvloeden op individuele verschillen in ADHD en daaraan gerelateerde gedragsproblemen.

1) Sekseverschillen in concentratie-problemen, hyperactiviteit en attention deficit hyperactivity disorder (ADHD)

Sekseverschillen in ADHD werden onderzocht in de hoofdstukken 3, 4 en 5. De nadruk van deze drie hoofdstukken lag vooral op de rol van het geslacht van het kind bij de beoordeling van de mate van zijn/haar gedragsproblemen door de ouder of leraar.

In hoofdstuk 3 heb ik onderzocht of de hogere prevalentie van ADHD in jongens het gevolg is van werkelijke verschillen in de mate van gedragsproblemen tussen jongens en meisjes. Een andere mogelijkheid is namelijk dat de prevalentie hoger is doordat de manier waarop de items van een meetinstrument ingevuld worden niet onafhankelijk is van het geslacht van het kind. Bijvoorbeeld, een item dat zou luiden: "gedraagt uw kind zich drukker dan andere kinderen wanneer hij/zij met auto's speelt" zal misschien eerder positief worden beantwoord door een leraar van een jongen met ADHD dan een leraar van een meisje van ADHD, ongeacht de mate van gedragsproblemen. Wanneer de score op een item afhangt van het geslacht van een kind, spreken we van schending van meetinvariantie. In hoofdstuk 3 heb ik de meetinvariantie onderzocht voor 4 schalen van de Conners Rating Scale: oppositioneel gedrag, cognitieve problemen-inattentie, hyperactiviteit en de ADHD-index. Deze vragenlijst werd ingevuld door de leraren van 1511 tweelingparen. Uit deze analyses bleek dat de 4 schalen meetinvariant zijn m.b.t. geslacht. Met andere woorden, de score op deze schalen wordt beinvloed door de mate van ADHD en niet door het geslacht van het kind. Uit dit onderzoek bleek verder dat de erfelijkheid van de 4 schalen (56-71%) gelijk is voor jongens en meisjes. Wel werd gevonden dat er specifieke genen zijn die alleen een rol spelen bij jongens of bij meisjes.

In hoofdstuk 4 onderzocht ik de invloed van het geslacht op de associatie tussen de Child Behavior Checklist (CBCL) score op concentratie-problemen en de Diagnostic and Statistical Manual of Mental Disorders (DSM) diagnose van ADHD. Van de kinderen met een lage CBCL score had 96% geen ADHD. Van de kinderen met een hoge CBCL score had 36% van de meisjes ADHD in vergelijking tot 59% van de jongens. De CBCL score voorspelde ADHD diagnose significant beter in jongens dan meisjes. Dit resultaat suggereert dat een meisje met concentratie-problemen minder snel gediagnosticeerd wordt dan een jongen met dezelfde mate van concentratie-problemen. Aangezien een besluit tot behandeling vaak genomen wordt op basis van een DSM-diagnose zou dit kunnen leiden tot onderbehandeling van meisjes met concentratie-problemen. Op deze mogelijke onderbehandeling van meisjes in vergelijking tot jongens werd dieper ingegaan in hoofdstuk 5.

Meer specifiek gesteld, hoofdstuk 5 concentreerde zich op het feit dat de ratio jongens:meisjes met ADHD veel hoger is in klinische populaties (9:1) dan in niet-klinische populaties (3:1). Deze discrepantie suggereert dat meisjes met ADHD minder vaak verwezen worden voor behandeling van hun klachten dan jongens met ADHD. Op basis van klinische interviews met 504 moeders van tweelingen constateerde ik dat meisjes met ADHD inderdaad minder vaak een klinische instelling (b.v. een kinderpsychiater of een RIAGG) bezoeken dan jongens met ADHD. Daarnaast wordt aan meisjes met ADHD minder vaak medicatie voorgeschreven dan aan jongens met ADHD. Een aantal mogelijke

verklaringen voor deze bevindingen werd onderzocht. Jongens en meisjes met ADHD bleken evenveel overige psychiatrische stoornissen te hebben en ze vertoonden evenveel leerproblemen op school (zoals blijven zitten en het volgen van speciaal onderwijs). Ook was de mate van concentratie-problemen en agressie hetzelfde volgens beoordelingen van de moeder. Beoordelingen van de leraar wezen echter uit dat jongens met ADHD op school meer concentratie-problemen en agressief gedrag vertonen dan meisjes met ADHD. We kunnen concluderen dat sekseverschillen afhankelijk zijn van de context waarin een kind zich bevindt: jongens met ADHD vertonen op school meer gedragsproblemen dan meisjes met ADHD maar thuis is dit verschil niet terug te vinden. Het feit dat jongens op school meer gedragsproblemen vertonen zou kunnen verklaren dat ze vaker doorverwezen worden voor behandeling.

2) HET ONTWIKKELEN VAN STATISTISCHE METHODEN VOOR DE ANALYSE VAN TWEELINGDATA

De analyse van gegevens die zijn verzameld in het kader van tweelingstudies van gedragsproblemen wordt gekenmerkt door een aantal interessante statistische problemen. Het doel van het simulatie-onderzoek in hoofdstuk 6 was om te onderzoeken wat het optimale selectie design is voor de situatie waarin een goedkope meting (X) is afgenomen in een grote, representatieve steekproef tweelingparen en een duurdere meting (Y) is afgenomen in een uit de steekproef geselecteerde groep. Deze simulatie is belangrijk voor mijn eigen studie waarin ik interview data heb verzameld in een deel van de totale tweelingpopulatie voor wie vragenlijsten beschikbaar waren. De scores op X worden gebruikt om vast te stellen welke tweelingparen het meest informatief zijn voor de detectie van genetische en omgevingsinvloeden op de variantie van Y. Zes verschillende selectie mogelijkheden werden vergeleken, bij elk van deze werd 12% van de totale steekproef geselecteerd. Ten eerste stelde ik vast dat de selectie van informatieve paren het meest effectief is wanneer de correlatie tussen X en Y hoog is. Verder was de afname in statistische power als gevolg van de selectie relatief klein. Een design waarin geselecteerd wordt op individueel niveau in plaats van op paarniveau is het beste selectiedesign. Het Extreem Concordant en Discordant Design (EDAC) heeft de voorkeur wanneer een additioneel doel van de studie is om specifieke genen te lokaliseren.

Gedragsproblemen worden vaak weergegeven aan de hand van sommatie van de scores op items. Deze somscores zijn over het algemeen niet normaal verdeeld. In hoofdstuk 7 onderzocht ik het effect van schending van normaliteit van de verdeling (scheefheid en kurtosis) op het schatten van genetische en omgevingsinvloeden met behulp van "normal theory Maximum Likelihood". Ik maakte een onderscheid tussen additieve genetische invloeden (A), niet-additieve genetische invloeden (D), gedeelde omgevingsinvloeden (C) en niet gedeelde omgevingsinvloeden (E). Wanneer een ACE model wordt gepast op niet-normaal verdeelde data wordt A correct geschat, C wordt onderschat en E wordt overschat. De bias bleef gelijk wanneer een worteltransformatie werd uitgevoerd op de scheef verdeelde data en bleef ook onveranderd bij het gebruik van een sterk normaliserende transformatie (normal score transformatie). Wanneer een ADE model wordt gepast op niet-normaal verdeelde data wordt A onderschat terwijl D en E worden overschat. In beide modellen werden de parameters correct geschat wanneer categorische data analyse werd gebruikt. De statistische power om A, D en C te detecteren nam echter wel af bij het gebruik van categorische data-analyse.

Een essentiele aanname in tweelingonderzoek is dat twee-eiige tweelingen net zo gelijk worden behandeld als een-eiige tweelingen. In hoofdstuk 8 verkende ik de mogelijkheden om in multivariate fenotypische data een schending van deze "Equal Environment Assumption" (EEA) te detecteren. De identificatie binnen dit model en de statistische power om schending van de EEA te detecteren werden onderzocht. Verder bestudeerde ik de mate van bias als gevolg van het negeren van schending van de EEA. De

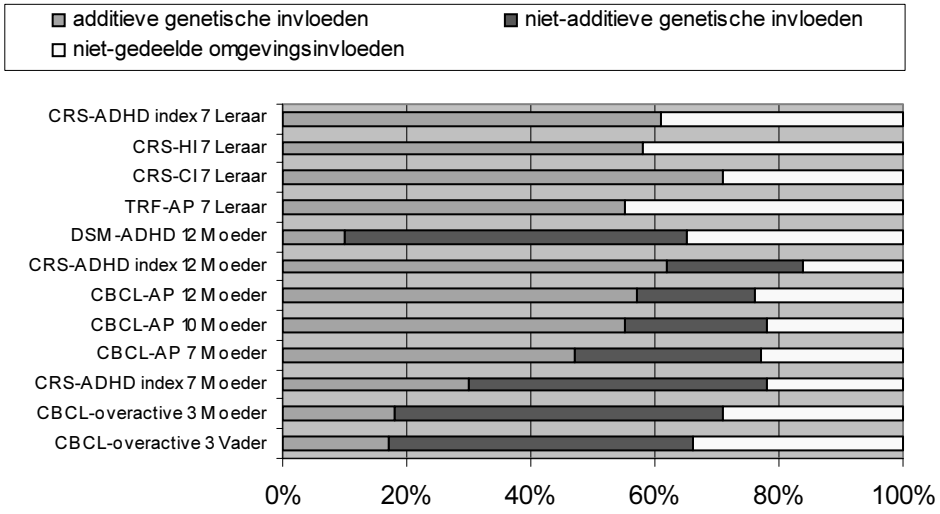
EEA kan zowel worden onderzocht in bivariate als in trivariate modellen. Het aantal tweelingparen dat nodig is om schending van de EEA te detecteren met een statische power van .80 (=.05) varieerde tussen 508 en 3576. De bias in de schattingen varieerde tussen de 5 en 34% voor A, en tussen de 4-34% voor C. De schatting van E werd niet beinvloed door schending van de EEA. Toepassing op empirische data liet zien dat de EEA niet geschonden wordt voor eigenschappen zoals spatiële vaardigheden en agressie.

3) Het onderzoeken van genetische en omgevingsinvloeden op individuele verschillen in ADHD en daaraan gerelateerde gedragsproblemen

In hoofdstukken 3, 9, 10 en 11 onderzocht ik de bijdragen van A, D en E tot individuele verschillen in concentratie-problemen, hyperactiviteit en ADHD. Hierbij maakte ik gebruik van data verkregen van verschillende informanten (moeder, vader en leraar) en met verschillende meetinstrumenten. De beoordeelde tweelingen varieerden in leeftijd. Figuur 1 geeft een samenvatting van de resultaten van dit onderzoek. Ook heb ik de resultaten van het onderzoek van Hudziak et al. (2005) toegevoegd aangezien in deze studie gebruik gemaakt werd van data van het Nederlandse Tweelingen Register.

Genetische invloeden verklaren het grootste deel van de individuele verschillen in beoordelingen door de ouders van de kinderen (66-84%). Niet-gedeelde omgevingsinvloeden verklaren de overige individuele verschillen. De erfelijkheid van beoordelingen door leerkrachten is ietwat lager en varieert tussen de 55 en 71%. Figuur 1 laat verder zien dat de totale genetische invloed (A+D) nauwelijks verschilt tussen de verschillende leeftijden. Bijvoorbeeld, de erfelijkheid van de CBCL-AP schaal ingevuld door de moeder is 77% op leeftijd 7, 78% op leeftijd 10 en 76% op leeftijd 12. De relatieve invloeden van de additieve en niet-additieve genetische invloeden varieren wel als een functie van leeftijd en instrument. Op leeftijd 3 bestaat de genetische invloed vooral uit niet-additieve effecten. Wanneer kinderen ouder worden neemt de invloed van de niet-additieve effecten af en neemt de invloed van de additieve effecten toe. De DSM-ADHD symptoom scores vormen een uitzondering. Ook op leeftijd 12 wordt de variatie in DSM-beoordelingen vooral verklaard door niet-additieve genetische effecten.

Figure 1. Figuur 1Het percentage variantie verklaard door additieve genetische effecten (A), niet-additieve genetische effecten (D) en niet-gedeelde omgevingsinvloeden (E) in 3, 7, 10 en 12-jarige kinderen



Note: CRS=Conners Rating Scale; TRF=Teacher Report Form; DSM=Diagnostic Statistical Manual of Mental Disorders; CBCL=Child Behavior Checklist; AP=concentratie-problemen, HI=hyperactiviteit, ADHD=attention deficit hyperactivity disorder, CI=cognitieve problemen-inattentie

In hoofdstuk 9 onderzocht ik de mate waarin genetische en omgevingsinvloeden afhankelijk zijn van de informant (vader/moeder) in een steekproef van 9689 3-jarige tweelingparen. Vijf CBCL-schalen werden onderzocht: agressief, oppositioneel, overactief, teruggetrokken en angstig/depressief gedrag. Individuele verschillen in de vader- en moederbeoordelingen worden deels beïnvloed door dezelfde aspecten van het gedrag van de kinderen maar ook voor een deel door aspecten die uniek zijn voor de beoordelingen van vaders en moeders. Individuele verschillen op alle 5 CBCL schalen worden gekenmerkt door relatief grote genetische invloeden. Gedeelde en niet-gedeelde omgevings invloeden zijn ook aanwezig, maar zijn minder belangrijk. Alleen overactief gedrag wordt niet beïnvloed door gedeelde omgevingsinvloeden.

In hoofdstuk 10 werden genetische en omgevings invloeden op individuele verschillen in moeder- en leraarbeoordelingen van concentratie-problemen onderzocht. Leraren vulden de Teacher Report Form (TRF) in (N=2259 paren) en moeders vulden de CBCL in (2057 paren) toen de kinderen 7 jaar oud waren. De correlaties van de leraarbeoordelingen waren hoger wanneer de 2 leden van een tweelingpaar door dezelfde leerkrachten werden beoordeeld dan wanneer de 2 leden door verschillende leraren werden beoordeeld. Dit kan enerzijds veroorzaakt worden doordat een leraar een bepaald idee heeft over welk soort gedrag afwijkend is en welk gedrag niet. Anderzijds kan de hogere correlatie ook het gevolg zijn van de invloed die een leraar heeft op het gedrag van de kinderen in zijn klas. Deze twee mogelijkheden kunnen met de huidige data en statistische modellen niet onderscheiden worden. Verder bleek dat 41% van de variatie in moeder- en leraar-beoordelingen het gevolg is van een gedeelde factor. De erfelijkheid van deze gedeelde factor is 78%. Verder zijn er ook beoordelaar-specifieke factoren. De erfelijkheid van deze factoren is 76% voor moeders en 39% voor leraren. Het feit dat de beoordelaar-specifieke factoren door genetische effecten beïnvloed worden laat zien dat moeders en leraren unieke aspecten van het gedrag van het kind beoordelen.

Het doel van hoofdstuk 11 was om te onderzoeken in welke mate individuele verschillen in drie verschillende instrumenten die AP en ADHD meten dezelfde genetische en omgevingsinvloeden reflecteren. Deze drie instrumenten werden afgenomen bij de moeders van tweelingen. De totale steekproef waarin tenminste 1 meting beschikbaar was, bestond uit 10916 kinderen afkomstig uit 5458 families. CBCL beoordelingen waren beschikbaar voor 10018, 6565 en 5780 kinderen op de leeftijden 7, 10 en 12 jaar. De Conners Rating Scale (4887 kinderen) en het DSM-interview (1006 kinderen) werden afgenomen op leeftijd 12. De fenotypische correlaties op leeftijd 12 varieerden tussen .45 tot .77. De varianties en covarianties werden vooral verklaard door genetische invloeden. De genetische correlaties van de metingen afgenomen op leeftijd 12 varieerden tussen de .61 en 1.00. Deze correlaties laten zien dat de genetische overlap tussen de verschillende instrumenten hoog is. Dit impliceert dat onderzoek naar de specifieke genen die een rol spelen bij ADHD gebruik kan maken van vragenlijst data en niet noodzakelijkerwijs van diagnostische interviews. Dit kan de kosten van zulke studies reduceren.

In hoofdstuk 12 werden de specifieke omgevingsinvloeden onderzocht welke mogelijk een rol spelen bij ADHD. Tweelingen werden vanaf de geboorte longitudinaal gevolgd. Vragenlijsten werden ingevuld door de moeder (op leeftijd 1, 2, 3, 7, 10 en 12 jaar van het kind), door de leraren (op leeftijd 7, 10 en 12 jaar van het kind) en door de tweelingen zelf (op leeftijd 12 jaar). Bij de moeders werd tevens een klinisch interview afgenomen. Uit deze grote longitudinale steekproef, selecteerden we discordante eeneiige tweelingparen, m.a.w., tweelingparen waarvan 1 lid van de tweeling hoog scoorde op ADHD en de andere laag. Deze werden gematched met concordante tweelingparen, m.a.w. tweelingparen waarbij beide kinderen hoog of juist laag scoorden. Kind-specifieke omgevingsvariabelen (b.v. duur in couveuse) werden vergeleken tussen de aangedane en niet-aangedane kinderen binnen de discordante paren. Omgevingsfactoren die gedeeld worden door de twee leden van een tweelingpaar (b.v. roken van de moeder tijdens de zwangerschap) werden vergeleken tussen de discordante paren en de concordante paren. De aangedane leden van de discordante tweelingparen scoorden hoger dan de niet-aangedane leden op een groot aantal metingen van concentratie-problemen, ADHD en andere gedragsproblemen zowel volgens de moeder, de leraar en de tweeling zelf. De aangedane leden hadden tevens een lager geboortegewicht en een slechtere vroege motorische ontwikkeling. Verschillen werden gevonden tussen de discordante groep, de concordant-hoge groep en de concordant-lage groep m.b.t. het roken door de moeder, het slapen in verschillende kamers, en het percentage 1-ouder gezinnen. Het roken door de moeder kwam meer voor in de concordant-hoge groep dan in de concordant-lage groep. De discordante tweelingen verschilden niet van beide groepen. Het slapen in verschillende kamers kwam meer voor bij discordante paren dan bij concordant-lage paren op de leeftijden 2-6 jaar. Meer concordant-hoge paren woonden in 1-ouder gezinnen dan concordant-lage paren of discordante paren.

REFERENCE LIST

Hudziak, J. J., Derks, E. M., Althoff, R. R., Rettew, D. C., & Boomsma, D. I. (2005). The genetic and environmental contributions to attention deficit hyperactivity disorder as measured by the Conners' Rating Scales - Revised. *American Journal of Psychiatry, 162,* 1614-1620.

LIST OF PUBLICATIONS

ARTICLES AND BOOK CHAPTERS

Derks EM, Dolan CV, and Boomsma DI (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research, 7,* 659-669.

Derks EM, Dolan, CV, and Boomsma, DI (2006). A test of the Equal Environment Assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics, 9,* 403-411.

Derks EM, Dolan CV, and Boomsma DI. Statistical power to detect genetic and environmental influences in the presence of data missing at random. *Twin Research and Human Genetics, in press.*

Derks EM, Dolan CV, Hudziak JJ, Neale MC, and Boomsma DI. Teacher Reports and the role of Sex and Genetics in Attention Deficit Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder (ODD). *Journal of Abnormal Psychology, in revision.*

Derks EM, Hudziak JJ, van Beijsterveldt CEM, Dolan CV, and Boomsma DI. Genetic analyses of maternal and teacher ratings on Attention Problems in seven-year-old Dutch twins. *Behavior Genetics, in press.*

Derks EM, Hudziak JJ, van Beijsterveldt CEM, Dolan CV, and Boomsma DI (2004). A study of genetic and environmental influences on maternal and paternal CBCL syndrome scores in a large sample of 3-year-old Dutch twins. *Behavior Genetics, 34,* 571-583.

Derks EM, Hudziak JJ, Dolan CV, Ferdinand RF, and Boomsma DI (2006). The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. *Comprehensive Psychiatry, 47,* 116-122.

Derks EM, Hudziak JJ, and Boomsma DI. Genetics of ADHD, Hyperactivity, and Attention Problems. In Y. Kim (Eds), *Handbook of Behavior Genetics, in press.*

Derks EM, Hudziak JJ , Dolan CV, van Beijsterveldt CEM, Verhulst FC, and Boomsma DI. Genetic and environmental influences on the relation between attention problems and attention deficit hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, submitted.*

Derks EM, Lehn H, Hudziak, JJ, Heutink P, van Beijsterveldt CEM, Boomsma DI. Attention problems and attention deficit hyperactivity disorder in discordant and concordant MZ twins: Evidence of environmental mediators. *Journal of the American Academy of Child and Adolescent Psychiatry, in press*

Althoff RR, Copeland WE, Stanger C, Derks EM, Todd RD, Neuman RJ, Boomsma DI, and Hudziak JJ. The Latent Class Structure of ADHD is Stable Across Informants. *Twin Research and Human Genetics, 4,* 507-522.

Boomsma DI, Rebollo I, Derks EM, Beijsterveldt CEM van, Althoff RR, Rettew DC, and Hudziak JJ. Longitudinal Stability of the CBCL-Juvenile Bipolar Disorder Phenotype: A Study in Dutch Twins. *Biological Psychiatry, in press.*

Bartels M, Derks EM, and Boomsma DI (2005). Oorzaken van individuele verschillen in normafwijkend gedrag. *Tijdschrift voor Criminologie, 47,* 129-135.

Boomsma DI, Vink JM, van Beijsterveldt CEM, Geus EJC, Beem AL, Mulder EJCM, Derks EM, Riese H, Willemsen AHM, Bartels M, van den Berg M, Kupper HM, Polderman JC, Posthuma D, Rietveld MJH, Stubbe JH, Knol LI, Stroet T and van Baal GCM (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Research, 5,* 1-6.

Hudziak JJ, Althoff RR, Derks EM, Faraone SV and Boomsma DI. Prevalence and genetic architecture of child behavior checklist-juvenile bipolar disorder (2005). *Biological Psychiatry, 58,* 562-568.

Hudziak JJ, van Beijsterveldt CEM, Bartels M, Rietveld MJH, Rettew DC, Derks EM, and Boomsma DI (2003). Individual differences in aggression: genetic analyses by age, gender, and informant in 3-, 7-, and 10-year-old Dutch twins. *Behavior Genetics, 33,* 575-589.

Hudziak JJ, Derks EM, Althoff RR, Rettew DC, and Boomsma DI (2005). The genetic and environmental contributions to attention deficit hyperactivity disorder as measured by the Conners' Rating Scales - Revised. *American Journal of Psychiatry, 162,* 1614-1620.

Hudziak JJ, Derks EM, Althoff RR, Copeland WM and Boomsma DI (2005). The genetic and environmental contributions to oppositional defiant behavior: A multi-informant twin study. *Journal of the American Academy of Child and Adolescent Psychiatry, 44,* 907-914.

Published Abstracts

Derks EM, Hudziak JJ, van Beijsterveldt CEM, and Boomsma DI. Genetic analyses of teacher ratings on aggression, attention problems and anxiety in 7-, 10-, and 12-year-old children (2004). *Behavior Genetics, 34,* 636-636.

Boomsma DI, van Beijsterveldt CEM, Derks EM, Bartels M, and Hudziak JJ (2004). Shared environmental factors involved in Anxiety/Depression during childhood: real or inflated by rater bias? A study from the Netherlands Twin Register. *Behavior Genetics, 34,* 632 -633.

Boomsma DI, van Beijsterveldt CEM, Derks EM, Bartels M, and Hudziak JJ. Longitudinal genetic analysis of Anxiety/Depression during childhood: A study from the Netherlands Twin Register (2004). *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 130B,* 23-23.

Boomsma DI, Derks EM, Heidema G, Vink JM, Dirkzwager HC, Hudziak JJ, Verhulst FC, Bartels M, Rietveld MJH, and van Beijsterveldt CEM (2002). Parental, teacher and self-reported aggression in Dutch twin families: Longitudinal studies in young and adult twins. *Behavior Genetics, 32,* 460-460.

Boomsma DI, van Beijsterveldt CEM, Derks EM, Bartels M, and Hudziak JJ. Longitudinal Genetic Analysis Of Anxiety/Depression During Childhood: A Study From The Netherlands Twin Register, XIIth World Congress of Psychiatric Genetics Annual Meeting, Dublin, Ireland October 2004. Published in Proceedings *World Congress of Psychiatry Genetics.*

Hudziak JJ, Derks EM, Althoff R, and Boomsma DI. Shared environmental contributions to the CBCL-Bipolar phenotype but not to the CBCL-AP phenotype in a sample of Dutch twins at ages 7, 10 and 12. American Academy of Child & Adolescent Psychiatry Association 50th Annual Meeting, Miami, FL, October 2003. *Published in Proceedings of American Academy of Child & Adolescent Psychiatry.*

DANKWOORD

Een proefschrift zou geen proefschrift zijn zonder dank te betuigen aan hen die dit verdienen. Graag wil ik beginnen met het bedanken van alle tweelingen, ouders en leraren die bereid waren om mee te doen aan het onderzoek van het Nederlands Tweelingen Register. Zonder hun medewerking zou het onderzoek waarop dit proefschrift gebaseerd is niet tot stand zijn gekomen. De schat aan informatie die dit onderzoek oplevert, zou niet half zo waardevol zijn zonder de hulp van de medewerkers van de administratie. Thérèse en collega's: bedankt voor de (letterlijk!) bergen werk die jullie verzetten. Toos, zonder jouw hulp zou ik eindeloos meer tijd kwijt zijn geweest aan het ordenen van de data. Wat een luxe positie om altijd de benodigde data van "de N" te kunnen halen. Daarnaast wil ik Natascha, Michiel en Jacqueline van het secretariaat bedanken voor al hun hulp bij allerlei praktische zaken.

Als onderdeel van mijn eigen onderzoek, zijn er interviewdata en DNA verzameld bij ruim 500 families. Gelukkig heb ik hiervoor de hulp gehad van een groot aantal mensen. Joost, bedankt voor het nauwkeurige werk dat je hebt verricht en voor de gezelligheid die je bracht. Ik zou niet weten wat ik zonder jouw hulp had gemoeten. Verder wil ik alle student-assistenten bedanken die de interviews hebben afgenomen.

Dit proefschrift zou niet tot stand zijn gekomen zonder de hulp van mijn drie begeleiders: prof. Boomsma, prof. Hudziak en dr. Dolan. Dorret, ik vraag me soms wel eens af of jouw dagen uit meer dan 24 uur bestaan. Ik vind het ten minste onbegrijpelijk hoe je het voor elkaar krijgt om altijd alle stukken die ik je opstuur snel te lezen, op de hoogte te zijn van de recente stand van zaken op heel veel verschillende gebieden en om een uitstekende begeleiding te bieden. Bedankt dat ik zoveel van je heb mogen leren. Jim, it was an honor to work with you. When I started my position as a Phd student, I thought that I would be mostly interested in the statistical modeling and much less in the phenotype. This has changed as a result of your enthusiasm for the subject, and your interest in the patients and their families. I hope that we can continue our collaboration in the future. Last, but not least, Conor, je bent nadat ik ongeveer een jaar als aio werkte betrokken geraakt bij mijn project. Wat ik van jou allemaal geleerd heb, is niet in woorden uit te drukken. De belangrijkste boodschap die je me hebt meegegeven is om nooit iets aan te nemen wat ik niet zelf heb onderzocht. Ook wil ik je bedanken voor de gezelligheid tijdens onze besprekingen, die overal konden plaatsvinden, bijvoorbeeld op de fiets naar Broek in Waterland.

I would like to thank the members of the reading committee: dr. Bartels, dr. Bochdanovits, prof. Kelderman, prof. Koot, prof. Neale and prof. Verhulst. Thank you for spending time on my thesis.

Het werken aan mijn proefschrift zou veel minder plezierig zijn geweest zonder de gezelligheid van mijn collega's. Omdat de lijst van namen te lang zou worden, en ik ongetwijfeld iemand zou vergeten, wil ik me beperken tot mijn kamergenoten: Tinca en Daniel. Wat hebben wij het gezellig gehad de afgelopen jaren! Het is prettig om je werk af en toe te kunnen onderbreken, en even te kletsen over niet-functiorende zonneschermen, de hypotheek, "het publiceren", enzovoorts.

Verder wil ik ook mijn paranimfen bedanken: Christel en Verena. Ik vind het een erg geruststellende gedachte dat jullie naast me zullen staan op de dag van mijn promotie. Christel, jouw humor leverde een grote bijdrage aan de sfeer op allerlei gelegenheden, waaronder de aio-etentjes. Het wordt nu echt tijd voor een post-doc etentje! Verena, we leerden elkaar kennen op de UVA, waar we allebei methodenleer studeerden. Daarna volgden etentjes, vakanties en heel veel gezelligheid. Ook nog bedankt voor het lenen van de Duitse Harry Potters. Helaas geldt hier ook dat begrijpen iets anders is dan toepassen, vandaar dat ik me toch maar tot het Nederlands beperkt heb.

Ook wil ik mijn familie en vrienden bedanken. Het zal voor jullie grotendeels een mysterie zijn gebleven wat ik hier precies doe. Juist dat werkte ontspannend, omdat ik daardoor het werk af en toe eens helemaal los kon laten. Mijn ouders wil ik in het bijzonder

bedanken omdat zij mij de mogelijkheid hebben gegeven om te gaan studeren. Ik ben blij dat ik daardoor werk kan doen waar ik echt plezier in heb!

Marc, jou wil ik ten eerste bedanken voor je heerlijke, gelijkmatige humeur. Het is maar goed dat we wat dat betreft heel verschillend zijn! Verder ben ik ontzettend blij met jou en ons busje. Ik hoop dat we nog heel veel weekendjes en vakanties samen weg kunnen, en vanaf volgend jaar met z'n drietjes!

Eske, september 2006