
JACKPOT


JACKPOT

A Genetic Study

of Problem Behaviors in Children

(een studie naar erfelijke invloeden op probleemgedrag bij kinderen)

PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Erasmus Universiteit Rotterdam

op gezag van de rector magnificus

Prof. Dr. C.J. Rijnvos

en volgens besluit van het college van dekanen.

de openbare verdediging zal plaats vinden op

woensdag 12 mei 1993 om 15.45 uur

door

Eduardus Johannes Cornelis Gerardus van den Oord

geboren te Oegstgeest

Promotie commissie

promotor:	Prof. Dr. F.C. Verhulst
co-promotor:	Dr. D.I. Boomsma
Overige leden:	Prof. Dr. H. Galjaard
	Prof. Dr. J.F. Orlebeke
	Prof. Dr. R.W. Trijsburg

This research was supported by a grant from the Sophia Foundation for Medical Research

Paranimfen: Drs. I. Lakeman

Drs. T. Stroet

acknowledgment/dankwoord

Ik zou iedereen willen bedanken die een bijdrage heeft geleverd aan de totstandkoming van dit proefschrift. In eerste instantie denk ik hierbij aan de ouders van tweelingen en adoptiekinderen die bereid waren de vragenlijsten in te vullen. Onmiskenbaar was ook de rol van mijn promotoren Dr. D.I. Boomsma en Prof. Dr. F.C. Verhulst. Zij hebben het project opgezet en begeleid. Mignon Roeleveld-Visser en Thérèse Stroet ben ik erkentelijk voor hun bijdrage aan het project. De samenwerking met Hans Koot liep ronduit gesmeerd. Jan van der Ende zou ik willen bedanken voor zijn meer dan royale collegialiteit. Dr. C.V. Dolan ben ik erkentelijk voor zijn bereidheid delen van het proefschrift te lezen op momenten dat dat nodig was. Mijn opleiding genoten bij Prof. G.J. Mellenbergh en Drs. F. Oort bleek van grote waarde te zijn. De leden van de kleine kommissie Prof. Dr. H. Galjaard, Prof. Dr. J.F. Orlebeke en Prof. Dr. R.W. Trijsburg, wil ik bedanken voor het lezen van het proefschrift. In het bijzonder denk ik hierbij aan Prof. Dr. J.F. Orlebeke daar hij ook in een eerdere fase al bij het onderzoek betrokken was.

Ontwerp: E. Regensburg

to Esther

Contents

1
Introduction

Aims of the study

Behavioral/emotional problems are common among children of preschool and school age. Verhulst, and Koot (1992, p. 130) reviewed prevalence studies published since 1965. They reported a median prevalence rate for general psychiatric dysfunction in children and adolescents of 13%. This number illustrates that problem behaviors in children present a public health problem that cannot be ignored.

During the 1960s and 1970s, many people came to dismiss the role of genetic factors in behavioral/emotional problems in children, and to emphasize the power of environmental influences (Rutter, 1991). However, recent years have shown an increased interest in the study of genetic factors (Plomin, in press; Rutter et al., 1990a). This has led to a broader recognition that genetic as well as environmental factors may be involved in children's problem behaviors.

Compared to the number of studies concerning the genetic influence in adult psychiatric disorders, only few have focussed on the role of genetic factors in child psychiatric conditions. A number of family, adoption, and twin studies have demonstrated the probable importance of genetic factors in relatively well-delineated child psychiatric conditions such as autism (Folstein, & Rutter, 1977), enuresis (Bakwin, 1971), tics (Pauls, Cohen, Heimbuch, Detlor, & Kidd, 1981), anorexia nervosa (Holland, Hall, Murray, Russel, & Crisp, 1984), and stuttering (Vandenberg, Singer, & Pauls, 1986). Other studies have investigated the genetic and environmental contributions to the commoner varieties of children's problem behaviors such as depression (Wierzbicki, 1987), hyperactivity (Goodman, & Stevenson, 1989a,b), delinquency (Rowe, 1983), and aggression (Ghodsian-Carpey, & Baker, 1987; Plomin, Foch, & Rowe, 1981). Nevertheless, for the vast majority of problem behaviors we cannot yet say whether genetic influence is significant, let alone estimate its magnitude (Plomin, in press).

The primary aim of this study was to address the most basic question of the extent of genetic involvement in the commoner varieties of problem behaviors in children. Estimates of genetic influences can be obtained by disentangling genetic and environmental influences. The genetic study in this dissertation is therefore as informative about environmental influences as it is about genetic influences. The value of a genetically informative design to study environmental influences is further illustrated by the possibility to assess the relative importance of two kinds of environmental influences. Environmental influences can be distinquished according to whether they have an impact on all children growing up in the same family, or uniquely influence one specific child. Parental rearing practices, illness/loss of a parent, or the socio-economic status are examples of possible shared environmental influences. Accidents, differential parental treatment, or peer group influences are examples of non-shared environmental influences because these are likely to affect the behavior of only the child of concern.

Disentangling genetic, shared environmental, and non-shared environmental influences may be scientifically and clinically useful. For instance, for most psychological characteristics in the area of personality, psychopathology, and cognition, the relevant environmental influences are not shared by children in the same family (Plomin, & Daniels, 1987). A possible implication is that research efforts and clinical interventions might

perhaps better focus on environmental variables that affect just one child, than on environmental variables that are assumed to affect all the children in the family.

An important part of this dissertation concerned problem behaviors in preschool children. Little is known about genetic influences on problem behaviors in preschool children, and the present study is one of the first reports on this subject. Further, a number of studies reported a substantial stability of problem behaviors in children (Richman, Stevenson, & Graham, 1982; Verhulst, & Van Der Ende, 1992a,b). Since early adjustment is an important predictor of the level of problem behavior at a later point in time, this argues for a greater understanding of the determinants of problem behaviors in young children. An increased knowledge might help to optimalize clinical interventions, and prevent later maladjustment.

Continuous variation

The majority of child psychiatric conditions do not fall into clearcut diagnostic categories (Verhulst, & Koot, 1992 p. 33). Problem behaviors in children generally involve quantitative variations of behavior that most children display to some degree. It is therefore preferable to examine the genetic influences on child psychiatric conditions assessed as quantitative variations of behavior rather than all-or-none categories.

From a genetic point of view it is likely that for these continuous variations the effects of many genes are involved (McGuffin, & Gottesman, 1985; Plomin, Rende, & Rutter, in press), and that methods of the quantitative genetic theory have to be applied for studying child psychiatric disorders. Quantitative genetics emerged in the early 1900s from disagreements between "Mendelians", who rediscovered Mendel's laws of inheritance, and so-called "biometricians" who felt that Mendel's laws derived from experiments with qualitative characteristics in pea plants were not applicable to complex characters in higher organisms (Mather, & Jinks, 1971 pp. 1-4; Plomin, 1986 p. 8). The resolution of the dispute came with the realization (Fisher, 1918) that Mendelian mechanisms of discrete inheritance apply to continuous variation too, but that the effects of many genes instead of the effect of a single gene are involved (McCleam, & DeFries, 1973 pp. 22-23). The simultaneous effect of many genes, each with a small effect, as well as the superimposition of truly continuous variation arising from non-genetic sources, causes continuous instead of discontinuous variation (Falconer, 1989 p. 104).

The Child Behavior Checklist

The clinical, medically oriented, tradition in psychiatry and the psychometric tradition in psychology are the two main approaches to assessment and taxonomy that have dominated research on child psychiatric disorders (Verhulst, & Koot, 1992 pp. 43-46). Within the medical tradition, disorders are classified as "present" versus "absent" by the use of a clinical interview. The psychometric approach to assessment typically uses quantitative ratings on scales which consist of sets of related items. Because of the continuous character of children's problem behaviors, the psychometric approach is likely to be more useful for a genetic study.

In (genetic) research, probably greater progress will be made on more narrowly defined areas of behavior rather than global diagnostic categories such as emotional problems or conduct disturbances (Plomin, Nitz, & Rowe, 1990; Vandenberg, Singer, & Pauls, 1986 p. 194). Narrowly defined syndromes such as hyperactivity, depression,

aggression may provide a better basis for detecting specific etiologies or predicting the outcome of specific treatments (Achenbach, & Edelbrock, 1984, p. 234).

In the present study, the Child Behavior Checklist (CBCL) was used for assessing problem behaviors in children. The CBCL is a widely used rating scale, developed by Achenbach (1966, 1978), and Achenbach and Edelbrock (1981, 1983) from the psychometric approach. It is probably the most elaborate studied assessment instrument in the area of child and adolescent psychopathology. Furthermore, the CBCL allows a distinction between a broad array of narrowly defined syndromes.

Structure of the dissertation

The dissertation consists of two parts. The first part, chapters 2, 3, and 4, comprises a review of genetic studies of problem behaviors in 4- to 18-year-old-children, and analyses on a sample of 11- to 15-year-old international adoptees. The second part, chapters 5, 6, 7 and 8, covers problem behaviors in children of preschool age. In this part results from analyses on a sample of 3-year-old twins are presented.

Part 1. A short introduction in the methods that have been applied for studying genetic influences on problem behaviors in children and adolescents, is presented in Chapter 2. This introduction is followed by a survey of findings from genetic studies of the commoner varieties of problem behaviors in children aged 4-18.

In preparation to the genetic analyses, the applicability of American CBCL/4-18 syndromes (Child Behavior Checklist for ages 4-18, Achenbach, 1991) in the sample of international adoptees was studied. Results are presented in chapter 3.

In chapter 4, the American CBCL/4-18 syndromes were used to study genetic and environmental influences on problem behaviors in the international adoptees. Twin data were used in the majority of genetic studies of problem behaviors in children. The adoption sample in the present study therefore provided an unique opportunity for a comparison with twin study inferences about genetic and environmental effects.

Part 2. For the genetic analyses on the sample of 3-year-old twins, the Child Behavior Checklist for Ages 2-3 (CBCL/2-3, Achenbach, 1992) was used to obtain parental ratings. About 60% of the items in the CBCL/2-3 have counterparts on the CBCL/4-18, while the remaining items have been developed specifically for ages 2-3.

In chapter 5, Dutch syndromes for the CBCL/2-3 were derived, to be used in the genetic analyses. The Dutch syndromes were obtained by performing item analyses on a clinical sample, a general population sample, and the twin sample from the present study.

In studying twin populations it is important to be able to generalize findings from the twin sample to the general population. To examine the representativeness of our twin sample, comparisons were made with a community sample consisting of 2-3-year-old singletons whose parents completed the CBCL/2-3. Results are presented in chapter 6.

Chapter 7 reports, for the separate syndrome scales, the genetic analyses on the parental ratings of problem behaviors in their 3-year-old twins. In chapter 8, genetic and environmental influences on covariances between the separate syndrome scales were studied. Such a multivariate genetic study is, for instance, useful to detect higher order syndromes that may be distinquished from other syndromes with respect to prognosis, course, or respons to clinical intervention.

In the final chapter, chapter 9, results from the genetic analyses were discussed. Attention was paid to the interpretation of the findings. Issues concerning use and misuse of genetic findings were addressed.

2
Literature Review

Introduction

This chapter reviews genetic studies of the commoner varieties of problem behaviors in children and adolescents. Possible relevant genetic studies of temperamental characteristics and personality features were also discussed. Because the first part of this dissertation concerns children aged 4 to 18, the mean age of the children in the selected genetic studies was also within these bounds.

First, genetic designs that have been used to study genetic influences on problem behaviors in children were briefly discussed. The other sections in this chapter review genetic studies of general psychiatric dysfunctioning in children, internalizing behaviors, antisocial and agressive behaviors, hyperactivity and attention problems. When multiple genetic designs had been used, family studies were reviewed first, then adoption studies, and finally twin studies. For each study the following information was reported: the design of the study; the assessment procedure or instrument; and the number and age of the subjects. The results for the total sample were reported, unless only results were available for girls and boys separately. The samples in most studies were too small to justify a further division.

Genetic designs

Twin studies

Twin data were used in most genetic studies of problem behaviors in children. In twin studies the difference in resemblance between monozygotic (MZ) twin pairs versus the resemblance between dizygotic (DZ) twin pairs is used to study genetic influences. MZ twins are genetically identical, DZ twins share only a proportion of their genetic information. A higher sibling resemblance between MZ twin pairs compared to DZ twin pairs is therefore suggestive of genetic influences.

The mathematics and assumptions of the classical twin design have been discussed by several authors (Plomin, DeFries, & McClearn, 1990; Falconer, 1989). Equation 1 shows the usual model for decomposing ratings of the child's behavior in a weighted sum of genetic, A, and environmental factors, C and E.

$$P_1 = hA_1 + cC_1 + eE_1$$
$$P_2 = hA_2 + cC_2 + eE_2 \ , \qquad (1)$$

with:

P is the observed behavior (phenotype),
A is the the unobserved additive genetic factor,
C is the unobserved shared environmental factor,
E is the unobserved non-shared environmental factor,
h, c, and e, are the loadings from P on respectively A, C, and E,
subscript 1 refers to the first twin, subscript 2 refers to the second twin,

and:

A, C and E do not correlate or interact.

A polygenic model assumes that observed behavior (the phenotype) is influenced by genetic information at several different loci on the chromosomes. A single locus consists of elements (alleles) that each can contribute to the observed scores. The additive genetic values, A, are simply the sum of the effects of the different alleles at one locus, as well as the sum of the effects across all loci that influence the character. The environmental contribution to a characteristic can be separated in environmental influences which are shared (C) and those that are not shared by siblings (E). Examples of shared environmental factors are socio-economic status or parental rearing practices. Illnesses, accidents, and differential parental treatment are examples of non-shared environmental influences. In equation 1 weights h (additive genetic effect), c (shared environmental effect), and e (effect of the non-shared environment) can be viewed as regression coefficients or factor loadings.

The extent to which a trait is determined by genetic influences is called the heritability of the trait. The heritability of a trait equals the genetic variance divided by the total variance. The total variance is the sum of the genetic, shared environmental, and non-shared environmental variance. In terms of model 1, the heritability can be expressed as: $h_s^2=h^2/h^2+c^2+e^2$ (subscript s refers to a standardized genetic effect). It can be interpreted as the proportion of genetic variance of the total variance. For instance, a heritability of .5 implies that 50% of the differences between subjects are accountable for in terms of genetic influences.

Estimates of the heritability, and the proportion environmental variance can be derived from differences in observed resemblance between MZ and DZ twin pairs. Figure 1 depicts the model presented in equation 1 for MZ and DZ twin pairs.

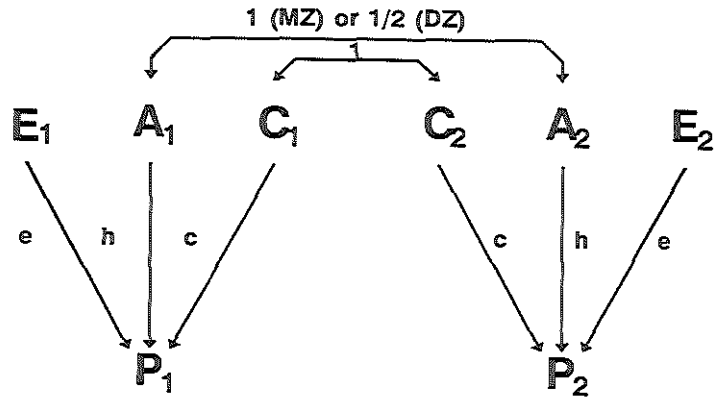


Figure 1. A model for twin resemblance.

For MZ twin pairs the correlation between the additive genetic factors of the twins equals 1, because MZ twins are genetically identical. DZ twins, who are genetically speaking full siblings, share only a part of their genetic information. The genetic correlation of DZ twins equals on the average .5 for DZ twins (Falconer 1989, p. 154).

For sibling pairs the correlation between the shared environmental factors equals by definition 1, and between non-shared environmental factors zero.

The additive genetic factor, and the shared environmental factor are identical for MZ twin pairs. Both these factors therefore contribute fully to the observed correlation between MZ twin pairs. Non-shared environmental influences are by definition unique for each separate twin, and do not contribute to twin pair resemblance. In terms of model 1, the observed correlation between MZ twins may therefore be expressed as (latent variables A, C, and E are scaled by equaling their variance to one): $r_{MZ}=h_s^2+c_s^2$ (subscript s refers to standardized effects). For DZ twins, the shared environmental influences are identical but the correlation between the additive genetic factors is only .5. The shared environmental factor contributes therefore fully to the DZ twin correlation, but the genetic factor only half. In terms of model 1, the correlation between DZ twins may therefore be expressed as (latent variables A, C, and E are scaled by equaling their variance to one): $r_{DZ}=.5h_s^2+c_s^2$.

Shared environmental factors are identical for MZ and DZ twin pairs. However, the correlation between the additive genetic factors of MZ twin pairs is twice the correlation between the additive genetic factors of DZ twins. The difference between the MZ versus DZ twin correlation has therefore to be doubled to estimate the heritability. This can be shown by substituting the expressions for the MZ and DZ twin correlations in this formula for estimating the heritability, and then to simplify the resulting term:

$$2(r_{MZ}-r_{DZ})=2\{(h_s^2+c_s^2)-(.5h_s^2+c_s^2)\}=2h_s^2+2c_s^2-h_s^2-2c_s^2=h_s^2.$$

In a similar way it can be shown that the proportion of shared environmental variance can be estimated by $c_s^2=2r_{DZ}-r_{MZ}$. Finally, the proportion non-shared environmental variance can be estimated by substraction $e_s^2=1-h_s^2-c_s^2$, which should correspond to $1-r_{MZ}$.

Care is needed when the above formulas are applied to twin data. Several studies suggested that the model in equation 1 may be too simplistic. For instance, in some areas of psychopathology evidence was found that persons with psychological complaints tend to choose partners with psychological complaints (Merikangas, Weissman, Prusoff, & John, 1988). A tendency of persons to mate non-randomly with regard to some characteristic is called assortative mating. In the presence of assortative mating, the formulas presented above do not yield accurate esimates of genetic and environmental contributions anymore.

When a model fitting approach is used, a test can be employed to judge the fit of the model. A chi-square test is often used for this purpose. A genetic model (e.g. model 1) is fitted to the data. When the model does not fit, the chi-square will be too large and the model is rejected. Using the rejected model, erroneous conclusions about the importance of genetic and environmental influences may be obtained. It is sometimes possible to fit a more appropriate model to the data, and to obtain correct estimates of the genetic and environmental effects. For instance, when psychopathology is assessed in parents as well as in their children, models may be fitted that account for assortative mating. The possibilty of these adaptions is one of the advantages of a model fitting approach. Compared to the formulas presented above, another advantage of a model fitting approach is that the estimates of genetic and environmental effects (h, c, e) are superior with respect to statistical criteria such as the precision with which genetic and environmental effects are estimated. Because of these advantages several studies in this review have used a model fitting approach. For these studies the proportion of genetic and environmental variance in the total variance was reported, instead of MZ or DZ twin correlations.

Some studies (e.g. Shields, 1977) in this review measured problem behaviors as

present versus absent. These studies reported concordance rates: the proportion of the total amount of twins pairs in which a disorder is present in both twins. Like with continuous data, a larger resemblance (concordance) between MZ twins compared to DZ twins implies genetic influences.

Adoption studies

When psychopathology is assessed in the biological parents of adopted children, groups of adoptees can be distinguished according to the mental health of their biological parents. This information about biological differences between groups of adoptees can be used to study genetic influences. A higher prevalence of psychopathology in adoptees with psychiatrically disturbed biological parents indicates genetic influences. The resemblance between the adoptive parent and the adoptive child provides an indication of the importance of the family environment.

Another adoption design which was used in one of the studies (Plomin, Defries, & Fulker, 1988) in this review, is called the sibling adoption design. In this design, the difference in resemblance between non-adopted biological siblings versus the resemblance between adopted children and the biological children of adoptive parents is used to study genetic influences. A larger sibling resemblance in the former group than in the latter group is indicative of genetic influences. Resemblance between the biological and adopted children in the family indicates shared environmental influences. The same model, as was discussed for twins, can be used in this sibling adoption design. However, the genetic correlation equals .5 in the group of biological siblings, and 0 in the group of adopted children and the biological children of adoptive parents. An estimate of the heritability can still be obtained by doubling the difference between the sibling correlations in both groups, but the sibling correlation in the group of adopted and biological children is itself a direct estimate of the proportion shared environmental variance.

Family studies

A number of studies in this review assessed the extent to which problem behaviors aggregate in families. In these studies familial aggregation is tested for by determining whether the incidence of the specified condition is greater in the relatives of disorderd subjects, than in the relatives of subjects without the disorder. A disadvantage of this design is that familial aggregation may result from either shared genes or environments. For psychiatric conditions that do not aggregate in families, it can be concluded that only non-shared/non familial environmental influences are important. However, when familial aggregation is found, it is not possible to separate the genetic from the shared/family environmental contribution to this loading. Family studies were therefore not reviewed in detail.

General psychiatric dysfunctioning in children

Cunningham, Cadoret, Loftus, and Edwards (1975) compared a group of 59 adoptees who were born of psychiatrically disturbed biological parents, with a comparison group of 54 adoptees who had psychiatrically normal biological parents. The adoptees in the comparison group were matched to the adoptees of the experimental group on sex, and the age of the biological mother at time of the birth. The median age at placement was about 3-4 months, the median age of the adoptees at the time of the interview was 17 years. Psychopathology in the biological parents was assessed by examining whether the adoption agency records mentioned psychiatric assessment or treatment for one or both parents. To assess psychopathology in the adoptees, each cooperating adopting family was

given a structured interview. Items were designed to elicit information relevant to behavior problems like: antisocial personality, hyperactivity, and neurotic symptoms such as anxiety and phobias. A psychiatrist, who did not know whether the child belonged to the experimental or control group, classified the recorded behavior of each adoptee in four categories: no, mild, moderate, or severe behavioral problems. Forty-six percent of the adoptees who were born of psychiatrically disturbed biological parents versus 31% of the adoptees in the comparison group had some degree of behavioral problems. In addition, 37% of the adoptees in the former group versus 14% of the adoptees in the latter group received or had received professional treatment for these problems. The concordance rates from this study suggested that a biological background of psychological complaints is predictive of problem behaviors, and that genetic factors are involved.

Evidence for genetic influence on a measure of general dysfunctioning in children was also found in a study by Edelbrock, Rende, and Plomin (1992). This study comprised 99 pairs MZ and 82 pairs of same-sex DZ twins (mean age 11.0 years). Ratings of twins' problem behaviors were obtained with the Child Behavior Checklist (CBCL). The CBCL is a widely used assessment instrument for assessing problem behaviors in children as reported by their parents (Achenbach, 1991). The total problem score is the sum of the scores for the 118 close ended items. For the total problem score Edelbrock, Rende, and Plomin (1992) found a correlation of .85 for MZ twins, and .66 for DZ twins. The formulas reported in section 2.1 suggested that genetic influences accounted for 38% (2x.85-2x.66), shared environmental influences for 47% (2x.66-.85=.47), and non-shared environmental influences for 15% (1-.38-.47=.15) of the total variance.

Graham, and Stevenson (1985) studied the genetic influence on behavioral problems, measured by the Rutter Parent and Teacher Scale. The sample consisted of 102 MZ, 111 same-sex DZ, and 72 opposite-sex DZ, 13-year-old twins. For respectively mother's, father's, and teacher's ratings, twin correlations were .60,.75,.74 for MZ twins, and .37,.44,.43 for same-sex DZ twins. The twin correlations from each rater suggested that genetic influences were more important than environmental influences.

Plomin, DeFries, and Fulker (1988, pp. 183-184) fitted several genetic models to parental ratings of children's behavioral problems on the CBCL. The sample consisted of 21 pairs adoptive, biologically unrelated siblings, and 16 pairs nonadoptive siblings. Children in this study were 4 years old. Results indicated that, for the total problem score of the CBCL, genetic influences were not important. However, the shared environmental influences were significant.

Shields (1977) reported results from three twin studies which used a broadly defined concept of behavior disorders. The first study (Rosanoff, Handy, & Plesset, 1941, in Shields, 1977) comprised 92 MZ, and 105 DZ same-sex pairs in which one or both twins were juvenile delinquents or had other childhood behavior difficulties like excessive shyness, excessive impulsiveness, or dull intelligence. The age of the twins was not reported. The concordance rate for MZ twins was 91%, for DZ twins 52%.

The second study investigated genetic influence on behavior disorders in a sample of 12 to 15 year-old-twins (Shields, 1954, in Shields, 1977). Assessments were mainly based on a history of the twins and descriptions of their behaviors on the lines of a semi-structured psychiatric social history. Observations of the twins' behavior and school reports were also taken into account. Shields rated the degree of psychopathology in the light of these data. Of the 62 same-sex twin pairs, 41 pairs had at some time in their lives been disturbed in their behaviors. The precise numbers of MZ and DZ twin pairs were not reported, but the concordance rate was 74% for MZ twins, and 50% for DZ twins.

Genetic influences appeared to have a relatively greater effect on the kind of childhood disorder than on its presence or severity. Of the 17 MZ twins who were concordant 14 had similar disorders, of the DZ twins only 1 of the 9 twins had a similar disorder.

The third study (Shields, 1977) also used a broadly defined concept of problem behavior, for example epilepsy, and reading disability were also assessed. The sample consisted of twins from same-sex pairs who were referred to a mental health institution. These twins were 16 years or younger. This is the only study in this review in which a clinical twin sample was used. The concordance rate for the 17 pairs of MZ twins was 65%, for the 24 pairs of DZ twins the concordance was 33%.

The MZ versus DZ differences in concordance rates in these three studies reported by Shields (1977), all implied genetic influences.

In conclusion, except for the small study by Plomin, DeFries, and Fulker (1988, pp. 183-184), all other studies suggested genetic influences. Marked was also the evidence for shared environmental influences found in the two CBCL studies (Plomin, DeFries, & Fulker, 1988, pp. 183-184; Edelbrock, Rende, & Plomin, 1992).

Internalizing behaviors

Much of the child clinical literature has focussed on two broad band groupings of problems (Achenbach, 1991a, p. 63). The first grouping of problem behaviors is covered in this section. This grouping is characterized by anxious, inhibited behavior. Over-controlled behavior or emotional disorder are other labels for this grouping which involves feelings of inferiority, self-consciousness, social withdrawal, shyness, anxiety, hypersensitivity, depression (Hersov, 1985). Somatic complaints such as headaches, stomachaches, or back pains, also frequently occur in combination with these characteristics (Last, 1989).

Internalizing grouping

Two twin studies with the CBCL internalizing scale showed both evidence for substantial genetic influence.

The study by Edelbrock, Rende, and Plomin (1992, see general psychiatric dysfunctioning in children) yielded MZ/DZ correlations of .75/50 for the CBCL internalizing scale.

Hewitt, Silberg, Neale, Eaves, and Erickson (1992) analyzed mothers' and fathers' ratings of the CBCL internalizing scale. Their sample was divided into an 8-11-year cohort and a 12-16-year cohort. The younger/older cohort consisted of 102/109 MZ female, 96/107 MZ male, 97/78 DZ female, 102/94 DZ male, and 103/95 opposite-sex DZ twin pairs. Because both parents rated each twin, a statistical model could be fitted that decomposed the total variance in unreliability, variance due to rater bias (defined as the tendency of an individual rater to over- or underestimate scores consistently), and trait variance. When not accounted for, unreliability of the assessment instrument spuriously inflates estimates of the non-shared environment and rater bias spuriously inflates estimates of the shared environment. In the younger cohort genetic, shared environmental, non-shared environmental influences accounted for 15%, 72%, 13% of the trait variance for girls, and 70%, 20%, 9% of the trait variance for boys. For the older cohort these percentages were 53%, 40%, 7% for girls, and 49%, 41%, 10% for boys.

Both CBCL studies indicated a heritability of about .50. The relatively small non-shared environmental component found in the study by Hewitt, Silberg, Neale, Eaves, and Erickson (1992) may be explained by the correction that is made for the unreliability

of the assessment instrument. However, compared to the CBCL study by Edelbrock, Rende, and Plomin (1992), the shared environmental influences were large despite the correction that was made for rater bias. Marked were also the sex differences found in the younger cohort in the study by Hewitt, Silberg, Neale, Eaves, and Erickson (1992).

Anxiety and depression

Several family studies showed that relatives, of children and adolescents with anxiety and/or unipolar affective disorders, have higher rates of such disorders too (Berg, 1976; Lavori, Keller, Beardslee, & Dorer, 1988; Livingston, Nugent, Rader, & Smith, 1985; Puig-Antich et al. 1989; Rosenbaum et al. 1988). Additionally, studies by Weissman et al.(1984; 1986) showed a marked increase in familial loading for depression when the age of onset in index cases was before the age of 20 years. The risk of depression before age 13 appeared also to be increased when parents themselves had an onset before age 20 (Weissman, Warner, Wickramatne, & Prusoff, 1988). All the family studies indicated that anxiety and depression aggregates in families. However, it is not possible to say whether this familial aggregation was caused by genetic or (shared) environmental influences.

Twin studies of anxiety and depression are summarized in Table 1. Self, parent, and teacher ratings were obtained in the study by Wierzbicki (1987). For self and parental ratings a modified version of the Beck Depression Inventory was used. For teacher ratings the Children's Depression Rating Scale was used. For instance, this latter questionnaire describes a variety of behaviors associated with depression like, depressed mood, weeping, low self esteem, social withdrawal, poor schoolwork, sleep/eating problems, physical complaints (Poznansky, Cook, & Carrol, 1979). Following this initial assessment, twins (using the depression scale of Wessman-Ricks Mood Scales) and their parents (using the Depression Adjective Check List) rated depressive symptomatology every evening for a period of 14 days. Only the teacher ratings on the Children's Depression Rating Scale showed no evidence of genetic influences.

The finding of no genetic influence as was reported for the teacher ratings in the study by Wierzbicki (1987) was also in contrast to other twin studies (see Table 1).

Edelbrock, Rende, and Plomin (1992) found evidence for genetic influence on the CBCL Anxious/Depressed scale.

Gottesman (1963) reported differences between MZ versus DZ correlations for self reported depression on the MMPI Depression scale. This scale contains sixty items. The items cover behaviors such as a lack of interest, apathy, and a denial of happiness or personal worth. Other items describe, a feeling of being incapable of performing work satisfactorily or controlling one's thought processes, physical symptoms, sleep disturbance, and lack of sociability.

Scarr (1966) studied anxiety in a sample which consisted entirely of girls. Maternal ratings on the Anxiety scale from Gough's Adjective Check List and observer ratings of anxiety using the Fels Child Behavior Scales were obtained. Both scales showed evidence for genetic influences

Stevenson, Batten, Cherner (1992) also found evidence for genetic influences. Their data consisted of self reports on the Revised Fear Survey Schedule for Children (FSSC-R). The total fear score is the sum of scores on 5 subscales: fear of failure; fear of the unknown; fear of injury and small animals; fear of danger; fear of medical procedures.

The personality trait 'neuroticism' is possibly related to anxiety and depression. For instance, the neuroticism scale from the Junior Eysenck Personality Inventory

Table 1 Twin studies of anxiety and depression.

Reference	Sample	Measure	twin correlations MZ	DZ
Edelbrock, Rende	99 MZ	CBCL Anxious/Depressed scale		
& Plomin (1992)	82 DZ		.64	.47
	11.0 years			
Gottesman (1963)	34 MZ	MMPI Depression scale		
	34 DZ		.47	.07
	16.2 years			
Scarr (1966)	24 MZ	Anxiety scale Gough's Adjective Checklist		
	28 DZ		.56	.03
	8.1 years	observer ratings of anxiety		
			.88	.28
Stevenson, Batten,	144 MZ	Revised Fear Survey Schedule for Children		
& Cherner (1992)	175 DZ		.78	.64
	11.8 years			
Wierzbicki (1987)	20 MZ	modified version Beck Depression Inventory		
	21 DZ	self rating	.53	.14
	6-16 years	Children's Depression Rating Scale		
		teacher rating	.54	.58
		Depression scale of Wessman-Ricks Mood Scales		
		self rating	.71	.15
		parent rating	.50	.11

contains questions like (Eysenk, 1969 pp. 266-269): do you ever feel 'just miserable' for no good reason; do you worry about awful things that might happen; do you often feel lonely; do you have many frightening dreams. Studies indicated that neuroticism not only applies to adults. Neuroticism can be assessed reliably in children too, even at the age of 4-5 years (Rachman, 1969). Genetic research suggested that about 50% of individual differences in neuroticism is accountable for in terms of hereditary influences (Eysenck, 1967 p. 210, Loehlin, 1989).

Genetic studies of neuroticism in children suggested that a heritability of the same magnitude. For instance, Newman, Freeman, and Holzinger (1966, p. 98) found MZ/DZ correlations of .56 and .37 for neuroticism (the Woodworth-Mathews Inventory), in a sample of 50 pairs MZ and 50 pairs DZ twins. Loelhin, and Nichols (1976) found MZ/DZ correlations of .48/.26 for girls and .58/.26 for boys. They used the Eysenck neuroticism scale from the California Psychological Inventory, to obtain ratings of neuroticism in a sample of 514 pairs MZ and 336 pairs DZ twins. Vandenberg (1962) using the neuroticism scale of Cattel's Junior Personality Quiz, reported a heritability of almost 70%(45 pairs MZ twins and 35 pairs DZ twins). Young, Eaves, and Eysenck (1980) used a model fitting approach to analyze data from the Junior Eysenck Personality Inventory. Data from 262 twin pairs and 182 singletons were analyzed in conjunction with adult EPQ data. An heritability estimate of .44 was obtained.

Family studies indicated a familial loading for anxiety and depression. Twin studies of anxiety and depression as well studies of neuroticism suggested that genetic influences make a larger contributions to this loading than shared environmental influences.

Social withdrawal

The study by Edelbrock, Rende, and Plomin (1992) showed evidence for genetic influences on the CBCL Withdrawn scale. These authors found MZ/DZ correlations of .53/.17.

Measurement scales for personality traits shyness and sociability frequently comprise behaviors such as shyness, timidity, and a preference to be alone. This indicates that genetic studies of shyness and sociability may be relevant to social withdrawal. Shyness (discomfort and inhibition that may occur in the presence of others) and sociability (need to be with people) are not merely opposing extremes of a bipolar dimension, but are better viewed as separate constructs (Cheek, & Buss, 1981).

Plomin, and Daniels reviewed genetic studies of shyness (1986), and sociability (1986). Eleven twin studies and 1 adoption study of shyness included children (mean age over 4 years) or adolescents. Most studies used questionnaire ratings. Five studies used a factor of Cattell's Sixteen Personality Factor Questionnaire. This factor has as its charateristic description (Plomin, & Daniels, 1986): shy, timid; restrained; threat-sensitive versus adventurous; 'thick-skinned'; socially bold. Behavior genetic data suggested that heredity influences individual differences in shyness perhaps more than in any other personality trait (Plomin, & Daniels, 1986). For instance, the mean MZ twin correlation for the 8 studies that reported it was .56 (for all 8 studies together there were 970 pairs), the mean DZ twin correlation was .10 (579 pairs).

Genetic influences on sociability appear to be substantial too (Plomin, 1986). Three studies in the review of Plomin (1986) used the sociability scale of the EASI questionnaire. This scale comprises the following items (Buss, & Plomin, 1975, p. 17): likes to be with others; makes friends easily; tends to be shy (reverse); tends to be independent (reverse); prefers to play by himself rather than with others (reverse). The other studies that were selected had scales wich showed resemblance to this scale. All, except one, studies used questionnaire ratings by parents. Subjects in six studies were in early childhood (mean age above 4 years), children in four studies were 6 to 10 years old, and subjects in one study were adolescents. Every study yielded evidence for substantial genetic influence. The mean correlation for the 7 studies that reported it was .63 for the total of 348 pair MZ twins, and .20 for the 278 DZ twins.

Genetic studies of shyness and sociability, as well as the study by Edelbrock, Rende, and Plomin (1992), suggested substantial genetic influences. However, low DZ twin correlations found in several studies suggested that the usual formulas to obtain heritability estimates may not be appropriate. The difference between MZ and DZ twin correlations implied genetic influences. However, in case of genetic influences, the correlation between DZ twins (who also share genetic information) has to be sufficiently large. For instance, by applying the usual formulas to obtain an estimate of the heritability in the study by Edelbrock, Rende, and Plomin (1992), a heritability of $2x(.53-.17)=.72$ is obtained. In the absence of shared environmental influences the DZ twin correlation equals half the heritability. The DZ twin correlation should therefore be something like .36 instead of .17 that was found. Sampling error can cause such impossible values. However, too low DZ twin correlations were also found for shyness, and sociability. This suggested that a model with only additive genetic, shared environmental, and non-shared

environmental influences may be too simplistic for behaviors associated with social withdrawal, and that heritability estimates based on this model may be incorrect.

Too low DZ twin correlation can occur for a number of reasons (see Goldsmith, 1989 for an enumeration). Plausible explanations include non-additive genetic effects such as interactions between alleles at the same locus (dominance) or interactions between alleles at different loci (epistasis). Parental expectations that MZ twins develop along more similar lines than DZ twins or a more similar treatment of MZ than of DZ twins, may inflate the MZ twin correlations and suggest too low DZ twin correlations. A final reason concerns sibling contrast or competition effects. For instance, twins might contrast their behaviors by trying to accentuate the existing differences between them. These contrast effects too, would result in too low DZ twin correlations and incorrect heritability estimates when not accounted for.

Somatic complaints

The revised version of Connors Parent Symptom Rating questionnaire was used in a twin study by O'Connor, Foch, Sherry, and Plomin (1980). This questionnaire contains a scale labeled 'Aches' that consists of two items: stomachaches; aches and pains. The sample consisted of 54 pairs of MZ twins, and 33 pairs of same-sex DZ twins. The mean age of the twins was 7.6 years (SD = 1.6 years). The twin correlation for the Aches scale was .70 for MZ twins, and .52 for DZ twins. This finding suggested almost equal parts of genetic, shared, and non-shared environmental influences.

Gottesman (1963, see anxiety and depression) using the MMPI hypochondriasis scale. The MMPI scale (Dahlstrom, & Welsh, 1960) contains 33 items, which describe generalized aches and pains, specific complaints about digestion, breathing, thinking, vision, and sleep as well as peculiarities of sensation. A few of the items relate to general health or competence. The MZ twin correlation for this scale was .39, the DZ twin correlation .21. Non-shared environmental influences were clearly larger than genetic influences. Shared environmental influences were very small.

Loehlin, and Nichols (1976, see anxiety and depression) reported twin correlations for a large number of items. For this review the follwing items were selected: nausea, headaches, and dizziness. The mean MZ/DZ twin correlations for these three items, pooled for girls and boys, were .27(514 MZ pairs)/.10(336 DZ pairs).

Edelbrock, Rende, and Plomin (1992) reported MZ/DZ correlations of .74/.35 for the CBCL Somatic Complaints scale. The 3 previous studies suggested a heritability between .3 and .4, the heritability in the study by Edelbrock, Rende, and Plomin (1992) was almost twice as large. However, the absence of shared environmental influences is in agreement with the studies by Gottesman (1963) and Loehlin, and Nichols (1976).

Antisocial and aggressive behaviors

A second group of problem behaviors which appears frequently in child clinical literature, is characterized by antisocial and aggressive behaviors. Conduct disorder, externalizing behavior, or undercontrolled behavior, are other terms for this grouping. It comprises delinquent behaviors such as theft or burglary, and other kinds of deviance, such as indiscipline, truancy, and physical aggression (West, 1985).

Externalizing grouping

Jary, and Stewart (1985) examined records, of 71 adopted children (mean age was 11.7) who were referred to a mental health institution. They selected adopted children who received a diagnosis of conduct disorder or attention deficit (with or without

hyperactivity), or whose presenting problems included aggressive behavior, disobedience, lying, stealing or any other antisocial behavior. The mean age at placement of the 37 adoptees with aggressive conduct disorder was 21.2 months. Detailed information on 34% of the biological parents was available for the initial group of 71 adoptees. The parents were diagnosed following the criteria of DSM-III. Precise numbers were not mentioned, but Jary, and Stewart reported that the biological mothers and fathers of the 37 adopted children diagnosed as having aggressive conduct disorder, had somewhat higher rates of psychiatric disorder than the corresponding parents of the adoptees with other diagnoses.

The two twin studies with the CBCL externalizing scale yielded also evidence of genetic influences.

Edelbrock, Rende, and Plomin (1992) reported for the CBCL externalizing scale MZ versus DZ twin correlations of .79 and .53.

Hewitt, Silberg, Neale, Eaves, and Erickson (1992, see internalizing problems in general) also reported results for the CBCL externalizing scale. In the younger cohort genetic, shared environmental, non-shared environmental influences accounted for 26%, 66%, 8% of the trait variance in girls, and 40%, 59%, 1% of the trait variance in boys. For the older cohort these percentages were 53%, 61%, 6% for girls, and 31%, 67%, 2% for boys. Like for the CBCL internalizing scale, the small non-shared environmental influences may be explained by the correction for the unreliabiliy of the scale. However, despite the correction for rater bias, the shared environmental influences were again large.

O'Connor, Foch, Sherry, and Plomin (1980, see section 3.3.2) also found evidence for genetic influences, but in contrast to the CBCL studies their results suggesetedmore modest shared environmental influences. These authors used the revised version of Connors Parent Symptom Rating questionnaire which contains a scale labeled Bullying. This scale has 6 items: bullying; hits or kicks other children; mean; sassy to grown-ups; fights constantly; picks on other children. MZ and DZ twin correlations were respectively .72 and .42.

In conclusion, genetic studies were suggestive of genetic influence. Remarkable was also the shared environmental component found for the CBCl externalizing scale (Edelbrock, Rende, and Plomin, 1992; Hewitt, Silberg, Neale, Eaves, and Erickson, 1992).

Antisocial and delinquent Behaviors

Little evidence for genetic influence on antisocial behaviors was found in a study of Bohman (1971, 1972). This study of 10-11-year-old children contained two groups which were especially useful for studying hereditary factors. The first group (n=168) consisted of children who, except one, were adopted within the first year of their life. The second group (n=124) consisted of children who were considered at birth difficult to place on account of retarded development, or somatic complications. More than half of the children in this group spent over nine months in institutions prior to placement in their adoptive or foster homes. These children were, at the time of the study, entirely separated from their biological environment.

Information on the biological fathers' social conduct was obtained by studying the registers concerning abuse of alcohol and crime. Registered criminality and alcohol abuse was considerably overrepresented among the biological fathers of both groups. However, for the first group of adopted children asocial symptoms (truancy, vagrancy, lying, stealing and pilfering, destructiveness) were reported relatively seldom by the adoptive parents or teachers, and were not reported more frequently in a comparison group of

same-sex class mates. This finding suggested that genetic influences were not important. Moreover, no relationship was found between the absence of presence of alcohol abuse and crime of the biological father, and a rating of overall adjustment in their children. In the second group of adoption/foster children a relation was found between the adjustment of girls, but not boys, and criminality in the biological fathers.

In a follow up (Bohman, & Sigvardsson, 1980) when the adopted children were 15 years old, the assessment of adjustment consisted of teacher ratings of: tension; withdrawal; aggressivness; psychomotor activity; ability to concentrate; contact with peers; social maturity (crimes, truancy, alcohol/drugs, running away); intelligence and school motivation. Also at this age no significant differences were found, in the adjustment of children whose biological fathers were registered for criminality and/or abuse of alcohol versus children whose fathers were not registered for criminality or abuse of alcohol. However, some differences in adjustment between children whose biological mothers were registered for criminality or alcohol abuse versus those whose mothers were not, were significant. The difference concerned psychomotor activity, and contact with peers in the first group of adopted children, and withdrawal, ability to concentrate, and school motivation in the second group of adoption/foster children.

Cadoret (1978, see also Cadoret, & Caine, 1980) selected adoptees who were separated at birth from their biological parent(s), and had no further contact with members of the biological family. The adoptees were placed in permanent adoptive homes. One group consisted of adoptees selected from backgrounds with a variety of psychopathology. Diagnoses of biological parents, and first and second degree family members were made by a psychiatrist on the basis of information from the adoption agency record, sometimes supplemented by hospital or court records. Another group of adoptees, who were matched for variables like sex and age to the adoptees of the first group, did not have a biological background with psychiatric conditions. The total sample consisted of 82 adults and 162 adolescents (aged 10-17).

A biological family background of alcohol abuse and antisocial behavior was associated with antisocial behaviors in the adoptees. These behaviors like, destructive, truant, fights, steals were contained in a structured interview given to the adoptive parents. However, a separate analysis on the child and adolescent sample (Cadoret, 1978) revealed no greater incidence of antisocial behavior in children and adolescents. The relationship between a biological family background of alcohol abuse and antisocial behavior and the antisocial behaviors in adoptees was caused by a significant correlation in the adult adoptee sample. For children and adolescents no evidence of genetic inluences was found. The authors did find a relation between psychopathology in adoptive parents or adoptive siblings, and the antisocial behaviors in the adoptees. This relation provided evidence for shared environmental influences.

Another study by Cadoret, Cain, and Crowe (1983) comprised, besides the sample mentioned above, two additional samples from other studies. The first sample consisted of a subsample of 40 adoptees from a study of Crowe (1974), and the second sample consisted of 108 adoptees from the study of Cadoret, Cunningham, Loftus, and Edwards (see also Cunningham, Cadoret, Loftus, and Edwards, 1975, section 2.2).

In both the study by Crowe (1974), and the the study by Cadoret, Cunningham, Loftus, and Edwards (1975) the same design had been used as in the study by Cadoret (1978). The adoptees from the studies by Crowe (1974), and Cadoret, Cunningham, Loftus, and Edwards (1975) were also separated at birth and matched to a group of control adoptees, and the same items concerning antisocial behaviors in the adoptee's

adolescent period were asked. However, the assessment of antisocial behavior in the adolescent period was based on reports of, at the time of the assessment, the adult adoptee him/herself for subjects from the study of Crowe (1974). For subjects from the study of Cadoret, Cunningham, Loftus, and Edwards (1975) assessment was based on ratings of the adoptive parents of antisocial behavior in the adoptee's adolesent period.

Genetic influence did not appear to be an important factor by itself, and was only important when also an adverse environmental factor was present (gene-environment interaction). A relation between psychopathology in adoptive parent or adoptive siblings, and antisocial behavior was found in all three samples. This suggested that shared environmental influences were important.

Edelbrock, Rende, and Plomin (1992, section 2.2) reported for the CBCL scale Delinquent Behavior MZ versus DZ twin correlations of .72 and .55.

Gottesman (1963, section anxiety and depression) reported results for the MMPI Psychopathic Deviate scale (Dahlstrom, & Welsh, 1960). This scale was developed to measure personality characteristics of antisocial persons. The 50 items of this scale describe a variety of behaviors. The items describe behaviors such as difficulties with authorities, poor morale, and sexual troubles. The twin correlation was .57 for MZ twins, and .18 for DZ twins.

Rowe (1983) used a model fitting approach to analyze self ratings of delinquent behaviors of 168 MZ and 97 DZ adolescent twin pairs. The measure consisted of 25 items that could be classified as theft, aggression, vandalism and minor delinquent act (e.g. trespassing, lying about age, causing a disturbance). Results from this study provided evidence for genetic influence. For girls MZ/DZ twin correlations were .62/.46, and for boys MZ/DZ correlations were .74/.52. Moreover, models without a genetic factor did not give an acceptable fit.

McGuffin and Gottesman (1985) pooled the findings of 6 twin studies concerning juvenile delinquency and crime. They found, for the total of 83 pairs of MZ twins and 61 DZ twins, a weighted average concordance rate of 87% for MZ twins and 72% for DZ twins. Their review suggested a small genetic component, and a large shared environmental component.

Studies of antisocial and delinquent behaviors indicated that genetic influences are present but are probably small in comparison to environmental influences. A number of adoption and twin studies suggested that not only non-shared environmental but also shared environmental influences may be important for antisocial behaviors (Cadoret, Cain, & Crowe, 1983; McGuffin, & Gottesman, 1985).

Aggression

Table 2 summarizes twin studies of aggression. Studies with the CBCL Aggression scale yielded evidence of substantial genetic influence (Edelbrock, Rende, and Plomin, 1992; Ghodsian-Carpey, and Baker, 1987).

Compared to the CBCL Aggression scale the observation checklist in the study by Ghodsian-Carpey, and Baker (1987) yielded smaller heritabilities. The observation checklist was based on a broad concept of aggression and included behaviors like destroying and damaging objects. Results for three measures were reported: the total number of behaviors checked from the 3 days the twins were observed by their mothers, the behaviors checked for the day both twins were observed together, and the behaviors checked for the two days each individual was observed separately.

Scarr (1966) also found a difference in MZ and DZ resemblance in her study. The mother ratings on the Aggression scale from Gough's Adjective Check List yielded a MZ

Table 2 Twin studies of aggression.

Reference	Sample	Measure	twin correlations MZ	DZ
Edelbrock, Rende	99 MZ	CBCL Aggression scale		
& Plomin (1992)	82 DZ		.75	.45
	11.0 years			
Ghodsian-Carpey,	21 MZ	CBCL Aggression scale		
& Baker (1987)	17 DZ		.78	.31
	5.2 years	observation checklist		
		twins obs. together	.70	.58
		twins obs. separately	.61	.27
		total	.65	.35
Owen,	18 MZ	projective test		
& sines (1970)	24 DZ		.09	.24
	6-14 years			
Plomin, Foch,	54 MZ	observations of children hitting a doll		
& Rowe (1981)	33 DZ	number of hits	.42	.42
	7.6 years	intensity of hits	.39	.47
		quadrant	.23	.41
Loehlin, & Nichols	504 MZ	three aggression items		
in Plomin, Foch,	328 DZ		.25	.17
& Rowe (1981)	high school twins			
Scarr (1966)	24 MZ	Aggression scale Cough's Adjective Checklist		
	28 DZ		.35	-.08
	8.1 years			
Vandenberg,	50 MZ	Aggression scale Stern's Activities Index		
in Plomin, Foch,	38 DZ	no MZ/DZ difference		
& Rowe (1980)	high school twins			

twin correlation of .35 and a DZ twin correlation of -.08.

In contrast to the three studies mentioned above other studies did not find evidence of genetic influences.

Owen and Sines (1970) found no evidence for genetic influences on a projective measure from the Missouri Children's Picture Series. Vandenberg (1967, in Plomin, Nitz, & Rowe, 1990) reported no significant genetic effect for the aggression scale of Stern's Activities Index.

Zero heritabilities were also found in the study of Plomin, Foch, and Rowe (1981). These authors videotaped twins hitting a 'bobo clown' doll (a 5-foot, inflated clown like figure, which is weighted at the bottom so that it rights itself after being knocked down). The number of hits, the intensity of hits, and the number of quadrants into which the child knocked the doll, were recorded from video tapes. In this study Plomin, Foch, and Rowe (1981) also reported twin correlations for three items (self-

ratings), from the study of Loehlin, and Nichols (1976, see 2.2), which are related to aggression: had a quarrel with a same-sex-friend; hit or slapped a same-sex person of your own age; and lost your temper. Twin correlations for these three items suggested only a small heritability.

The results from the three studies of aggression in childhood and the four studies in adolescence did not show a consistent pattern. Heritability estimates ranged from zero (Plomin, Foch, & Rowe, 1980) to .93 (Ghodsian-Carpey, & Baker, 1987).

Hyperactivity and Attention Problems

The disorder covered in this section is characterized by a disorganized and chaotic style of behavior, including restlessness and inattention. Terms like 'minimal cerebral dysfunction' and 'minimal brain damage' were also used to describe this disorder (Cantwell, 1975). It also matches the diagnosis of Attention Deficit Hyperactivity Disorder in DSM-III-R (American Psychiatric Association, 1987 p. 50).

Several studies suggested a familial loading for hyperactivity (Biederman et al. 1986; Cantwell, 1972; Morrison, & Stewart, 1971; Welner, Welner, Stewart, Palkes, & Wish, 1977). These studies compared biological first, and in some cases, second degree relatives with the relatives of normal controls. The biological parents of hyperactive children showed increased prevalence rates for alcoholism, sociopathy, and hysteria. Interviews with parents also indicated that hyperactivity occurred, or had occurred during childhood, more often in the biological first and second degree relatives of hyperactive children than in the relatives of non-hyperactive controls. Three other studies (Cantwell, 1975; Deutsch, 1990; Morris, & Stewart, 1973) included an additional group of adoptive hyperactive children and their adoptive relatives. The frequency of psychopathology/hyperactivity was less in the adoptive relatives of hyperactive adoptees than in the biological relatives of hyperactive non-adopted children, and resembled the group of normal controls.

Cadoret, Cunningham, Loftus, and Edwards (1975) studied hyperactivity in the same sample as was used in the study of Cunningham, Cadoret, Loftus, and Edwards (1975, see section 2.2). Answers of the adoptive parents to 3 hyperactive items were used to assess hyperactivity in adoptees. These items were: is he/she active during quiet periods or can he/she rest and lie quietly during quiet periods; when involved in an activity can he/she concentrate for ten minutes or more (reverse); when involved in an activity can his/her attention be easily diverted. For boys, but not for girls, the number of traits composing the hyperactive syndrome were significantly higher in the experimental than in the control group. This finding suggested genetic influences on hyperactivity in boys, but not in girls.

Safer (1973) screened a group of foster children referred to a mental health institution for 'minimal brain dysfunction' (MBD). Seventeen children were selected because they met the following criteria: a diagnosis of MBD by the examining psychiatrist or psychologist, defects in learning, attention, and behavior characteristic of the clinical picture, no evidence of organic cerebral insult, no coexistent diagnosis of psychosis, and an IQ over 70. Medical and social service charts, of 19 siblings and 22 half siblings of these 17 selected children, were examined by three raters. The full and half siblings, who were 5 to 9 years old when they themselves were adopted, were raised apart from the 17 children with MBD. Nine of the 19 siblings versus 5 of the 22 half siblings were found to be hyperactive, and 9 of the 19 siblings versus 3 of the 22 half siblings had a short

attention span. This finding suggested genetic influences.

Table 3 summarizes twin studies relevant to hyperactivity and attention problems. Most of the studies in Table 3 used questionnaire ratings. The studies of Buss, and Plomin (1975, p. 19) and Plomin (in Plomin, 1986 p. 214) used the EASI activity scale. This scale consists of 5 items (Buss, & Plomin 1975, p. 17): Child is always on the go; Child likes to be off and running as soon he wakes up in the morning; child cannot sit still long; Child prefers quiet games such as block play or coloring to more active games (reverse); child fidgets at meals and similar occasions. Goodman, and Stevenson (1989a,b) studied hyperactivity in the same sample as reported in the study of Graham, and Stevenson (1985, see general psychiatric dysfunctioning in children). Three hyperactivity items (squirmy; restless; cannot settle) from the Rutter Parent and Teacher Scale were used. In their study, Matheny, and Dolan (1980) used a two item scale to assess activity in 7 to 10 year old same-sex twins: overly active; and inattentive. The revised Connors Parent Symptom Rating questionnaire was used in the study of O'Connor, Foch, Sherry, & Plomin (1981, see somatic complaints). This questionnaire contains a scale labeled Tense and a scale labeled Restless. It has been shown that these scales, as well as three other scales from this questionnaire, can distinguish between hyperactive and non-hyperactive children (O' Connor, Foch, Sherry, & Plomin, 1981). The following six items constitute Tense: gets stiff and rigid; twitches/jerks; throws himself around; shakes; chews on clothes/blankets or others; picks at things such as hair/clothing etc. Restless contains the following items: restless; can't keep still; always into things; blames others for his mistakes. Willerman (1973) used the Activity Level Questionnaire developed by Werry, Weiss, and Peters. This questionnaire contains 32 items distributed over behaviors at mealtime, while watching television, doing homework, playing, sleeping, away from home (except school), and at school. The items describe behaviors such as, child talks excessively, wiggles, manipulates objects or body, inability for quiet play, restlessness, and interrupts.

Two studies in Table 3 used observer ratings. The study by Goldsmith, and Gottesman (1981) was a longitudional study. In their study ratings were made by trained psychologists during mental and motor testing and during free play. In the study of Torgersen (1982), raters scored tape-recordings of open-ended semi-structured interviews given to the twins' mothers.

The sample from the study by O'Connor, Foch, Sherry, & Plomin (1981) was also used in the study by Plomin, and Foch (1980). This study included three observational measures which may be relevant to hyperactivity: activity, fidgeting, and selective attention. Activity was assessed by means of a pedometer worn at the waist to record up and down movements of the trunk. Fidgeting was measured by video tape analysis of a 9-minute "rest" period during which the child was asked to lie in a beanbag chair as quiet as possible. Selective attention was measured by an auditory test. The child wore earphones and heard a tape recording of words such as "shoe". The child's task was to point to the picture that represented the word on a card containing four pictures. After a practice section to ensure that the child knew the words and the pictures, 11 words were presented with no background noise. The next two phases of the test involved increasing background noise. the test is called selective attention, because it measures an individual's ability to attend to a listening task in the presence of competing noise.

The majority of the family, adoption, and twin studies suggested genetic influences. Twin studies suggested a large heritability. However, except for the studies by Willerman (1973) and Plomin, and Foch (1980), too low DZ twin correlations were

Table 3 Twin studies relevant to hyperactivity and attention problems.

Reference	Sample	Measure	twin correlations MZ	DZ
Buss,	81 MZ	EASI activity scale	.62	.09
& Plomin (1975)	57 DZ			
	4.6 years			
Edelbrock, Rende,	99 MZ	CBCL Attention Problems		
& Plomin (1992)	82 DZ		.68	.29
	11.0 years			
Goldsmith,	189 MZ	Observer ratings		
& Gottesman (1981)	315 DZ	difference between MZ and DZ cor.		
	twins 4 years	activity	.10	
		attention span	.31	
	twins 7 years	activity	.11	
		attention span	-.16	
Goodman,	102 MZ	Three items Rutter Parent/Teacher scale		
& Stevenson (1989b)	111 DZ	Mothers' ratings	.68	-.08
	all twins 13 years	Fathers' ratings	.48	.21
		Teachers' ratings	.62	.26
Matheny,	68 MZ	Activity/	.66	.19
& Dolan (1980)	37 DZ	Distractability		
	median 8 years			
O'Connor, Foch, Sherry	54 MZ	revised CPSR		
& Plomin (1980)	33 DZ	Tense	.84	.15
	7.6 years	Restless	.70	.26
Plomin				
(In Plomin 1986)	51 MZ	EASI activity scale	.73	.05
	33 DZ			
	7.6 years			
Plomin,	51 MZ	pedometer	.99	.94
& Foch (1980)	32 DZ	fidgeting	.95	.51
	7.6 years	selective attention	.42	.50
Torgersen (1982)	34 MZ	open-ended semi-structured interview		
	16 DZ	activity	.93	.14
	all twins 6 years	attention span/	.73	-.27
		persistence		
Willerman (1973)	54 MZ	activity	.88	.59
	39 DZ			
	4.2 years			

found. This indicated that a model with only additive genetic, shared environmental, and non-shared environmental influences may be too simplistic and that heritability estimates obtained from this model may be incorrect for behaviors associated with (hyper)activity.

The DZ twin correlations for the behavioral measures used in the study of Plomin, and Foch (1980) were consistent with a model that allows additive genetic, shared environmental, and non-shared environmental influences. However, results for the three measures in this study showed an inconsistent pattern. The heritabilities ranged from almost zero for "selective attention" and the " pedometer" too almost .90 for "fidgeting". The non-shared environmental influences ranged from almost zero for the "pedometer" to about 50% for "selective attention". The shared environmental influences ranged from hardly 10% for "fidgeting" to almost 90% for the "pedometer".

Conclusions

For a number of reasons, it is difficult to draw firm conclusions concerning the importance of genetic and environmental influences on specific problem behaviors.

Firstly, the sample sizes in most studies were small. This may explain inconsistent findings, as for instance were found for aggressive behavior.

Secondly, a variety of assessment procedures and instruments were used. It is therefore possible that, although the same labels were used, different studies addressed different problem behaviors. Further, in a number of studies assessment procedures were used for which the validity and reliability was not clearly established. In these cases it was difficult to assess what was measured, how well it was measured, and consequently what meaning could be attached to the results.

Thirdly, in most twin studies models with additive genetic, shared environmental, and non-shared environmental influences were used, to obtain estimates of genetic and environmental influences. However, only in a few cases a model fitting approach was used to test the applicability of this model. With an invalid model, incorrect estimates of genetic and environmental influences may be obtained. For instance, for behaviors associated with social withdrawal and (hyper)activity, the too low DZ twin correlations found in most studies suggested that models with additive genetic, shared environmental, and non-shared environmental influences may be too simplistic. For these scales, estimates of genetic and environmental influences based on this model are likely to be incorrect.

Finally, estimates of genetic and environmental influences apply only to a particular population and its environmental circumstances at the time of the study (Rutter et al. 1991). Inconsistent findings may therefore reflect differences between populations.

To the extent that it is possible to draw general conclusions, it appeared that genetic influences were important to most problem behaviors. For psychological characteristics in the area of personality, psychopathology, and cognition, non-shared environmental influences are often more important than genetic, and shared environmental influences (Plomin, & Daniels, 1987). Shared environmental influences are often least important. Marked was therefore the evidence for shared environmental influences on antisocial behaviors.

3

Psychometric Properties of Achenbach's Cross-Informant Syndrome Constructs in a Sample of International Adoptees

Edwin J.C.G. Van Den Oord[1,2], Frank C. Verhulst[1], and Dorret I. Boomsma[2].

Abstract

To facilitate the coordination of questionnaire ratings from different informants assessing psychopathology in children and adolescents, Achenbach (1991a) derived so called cross-informant syndrome constructs. The validity of the cross-informant syndrome constructs and the content validity of Child Behavior Checklist items were studied in a Dutch sample of international adoptees (N=2,148). Results were cross-validated on a clinical sample (N=1,387). Support was found for the validity of the cross-informant syndrome constructs. In the sample from the present study, the contribution of a number of items to the scales of the cross-informant syndrome constructs was questionable. These items had very low variances, were not indicators of just one construct, or did not improve the reliability of the scale.

Introduction

The Child Behavior Checklist (CBCL) is a widely used rating scale (developed by Achenbach, 1966,1978, and Achenbach, & Edelbrock 1981,1983), for assessing problem behaviors and competencies in children and adolescents as reported by their parents. The Teacher's Report Form (TRF) and the Youth Self-Report (YSR) were derived from the CBCL to obtain reports from teachers and adolescents themselves. Different informants, such as parents, teachers or clinicians, seeing the child or adolescent under different conditions often disagree on the presence and severity of problem behavior (Achenbach, McConaughy, & Howell, 1987). This disagreement should not automatically be regarded as error. Instead, each informant may, from his own perspective, provide valid information on the children's functioning (Verhulst & Van Der Ende, 1991).

To facilitate the coordination of ratings on the CBCL, TRF, and YSR, Achenbach (1991a) derived so called cross-informant syndrome constructs. Several steps were taken to derive these constructs (Achenbach, 1991a, pp. 6-7). For each sex/age group, principal components analyses were performed on (a) all the problem items, and (b) problem items common to the CBCL, YSR and TRF. The syndromes obtained from these analyses were compared across sex/age groups, to identify syndromes that were similar for multiple groups. A core syndrome was derived from items that were common to the versions of the syndrome for most sex/age groups. Finally, for each core syndrome having counterparts in at least two of the three instruments, a cross-informant syndrome construct was

[1] Department of Child and Adolescent Psychiatry, Sophia Children's Hospital-Erasmus University Rotterdam, the Netherlands

[2] Department of Psychonomics, Free University, Amsterdam, the Netherlands

derived from items that were common to the core syndromes for at least two of the instruments. The cross-informant syndrome constructs were labeled Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior. Each construct is measured by a scale which consists of several items. Although the constructs are similar for each informant, some items in the scales are specific to the CBCL, YSR, or TRF. Both, constructs and scales, are identical for girls and boys and age groups 4-11 and 12-18.

The cross-informant syndrome constructs offer a number of advantages in comparison to earlier reported syndromes that are specific for each sex and age groups 6-11 and 12-16. Not only is the coordination of data from different informants facilitated. Sex and age differences are also studied more easily because each group is rated on the same scale. It was therefore decided to use these constructs in research with CBCL data of international adoptees living in the Netherlands.

In the present paper, the validity of the cross-informant syndrome constructs and the content validity of Child Behavior Checklist items were studied in the sample of international adoptees. Adopted children may show an increased genetic vulnerability (Verhulst et al., 1990b; Rutter et al., 1990a), and often have experienced more negative environmental influences (discontinuous caretaking, deprivation/abuse, malnutrition and medical conditions) which may put them at elevated risk for maladjustment (Verhulst, 1992). On the other hand, the selection of 'suitable' adoptive homes may counteract some of the negative effects of the early adverse environments (Tizard, 1977; Verhulst, 1992). These unusual characteristics may affect the psychometric properties of the cross-informant syndrome constructs.

There were also two more general points. Firstly, Achenbach (1991b) found support for the validity of the cross-informant syndrome constructs by comparing them with syndromes derived from other instruments. However, the psychometric properties of the cross-informant syndrome constructs have not yet been studied extensively in clinical or in community samples. For instance, factor loadings have not been reported for the CBCL, TRF, or the YSR. These loadings are important for studying the content validity of items. Secondly, little is known about the applicability of the American cross-informant constructs in Dutch samples.

The aim of the present paper was to establish the properties of the American factor structure in a sample of international adoptees. Confirmatory factor analysis was therefore used, instead of exploratory factor analysis. In confirmatory factor analysis a factor model has to be specified in advance. By specifying a model based on the findings of Achenbach (1991a), the cross-informant syndrome constructs were studied in the sample of international adoptees. The use of a factor model was justifiable here, because cross-informant syndrome constructs are to be viewed as hypothetical abstractions or, in statistical language, 'latent variables' (Achenbach, 1991a p. 44). The validity of the constructs was studied by evaluating the factor model. Factor loadings were used to assess the content validity of the items.

Method

Assessment instrument

The Child Behavior Checklist for children aged 4 to 18 (CBCL/4-18) is a rating scale for assessing problem behaviors and competencies in children and adolescents as reported by one of their parents. It consists of 20 competence items, not part of the cross-

informant syndrome constructs, and 120 problem items. The CBCL was translated into Dutch with the help of a linguist. The scales derived by Achenbach (1991a) comprise 85 problem items. Parents are requested to circle a 0 if the problem is not true of a child, a 1 if the item is somewhat or sometimes true, and a 2 if it is very true or often true.

Model

The mainframe version of the computer program LISREL 7 (Jöreskog, & Sörbom, 1989) was used to fit Jöreskog's (1971) congeneric measurement model (1) to the data. When observed and latent variables are expressed as deviations from their means, the congeneric model can be written as (using LISREL notation):

$$x_i = \sum_{j=1}^{n} \lambda_{ij}\xi_j + \delta_i \quad , \qquad (1)$$

with:

x_i is the i^{th} item, λ_{ij} is the loading from the i^{th} item on the j^{th} factor, ξ_j is the j^{th} factor, δ_i is the error of measurement of the i^{th} item, i=1...h with h i s the number of items, j=1...n with n is the number of factors,

and assumptions:

Factors and measurement errors are uncorrelated, the mean measurement error is zero.

The congeneric measurement model assumes that item scores consist of scores on underlying factors and errors of measurement. The congeneric model allows each item to have different factor loadings and error variances.

A model for the Cross-informant syndrome constructs

An initial model was based on the scales of the cross-informant syndrome constructs as reported by Achenbach (1991a) for the CBCL. A loading from an item on a construct was estimated when the item occurred in the scale of the construct. Otherwise the loading was fixed at zero. All correlations between the cross-informant syndrome constructs were estimated.

It was examined whether this initial model had to be elaborated with additional parameters, because its pattern of factor loadings was very restrictive and it did not allow correlated errors of measurement. These respecifications were guided by statistics (MI, EPC) LISREL can provide. The modification index (MI) gives an estimate of the improvement in fit of the model (using the chi-square statistic), in case a fixed parameter would have been left free. The EPC (expected parameter change) gives an estimate of a parameter in case that fixed parameter had been estimated instead.

Parameters with the highest MI's and EPC's were released. Then again the model was fitted to the data MI's and EPC's were inspected to see if more additional parameters had to be estimated. Every time a model was fitted to the data it was also examined if it could be simplified. Factor loadings (standardized) less than .125 and error of measurement correlations less than .05 were fixed at zero the next time the model was fitted. These values were considered non-substantial. This procedure of estimating new parameters and fixing non-substantial parameters continued until the overall fit indices did not improve anymore.

Polychoric correlations were used as input. Response scales of CBCL items are ordinal. Ordinary product moment correlations assume continuous response scales. Analyzing these correlations results in underestimates of factor loadings and in overestimates of unique variances (Jöreskog, & Sörbom, 1988 p. 1-16). Polychoric correlations

are appropriate for items with ordinal response scales. They can be viewed as estimates of the correlations between the items in case their scales had been continuous. Polychoric correlations were computed with PRELIS (Jöreskog, & Sörbom, 1988), under assumption of bivariate normality of the underlying continuous variables. PRELIS is a preprocessor of LISREL.

Cronbach's alpha is frequently used as a reliability/internal consistency coefficient. Bollen (1989, p. 216) showed that Cronbach's alpha is only appropriate for more restrictive measurement models (e.g. assuming that factor loadings of all items are equal). Bollen (1989, p. 221) also indicated that this coefficient does not make an allowance for correlated errors of measurement, and is not appropriate for items which are influenced by multiple factors. The reliability of a set of congeneric items can be computed with the formula given by Jöreskog (1971). Because this formula applies to models with one factor, it was adjusted to the case of multiple factors. Like in Cronbach's alpha, reliability equals the proportion of variance scale H has in common with the factor ξ_k it is supposed to measure. When latent variables are scaled by equaling their variance to one, then:

$$\rho^2_{\xi_k H} = \frac{\left[\sum_{i=1}^{h}\sum_{j=1}^{n} \lambda_{ij} COV(\xi_k,\xi_j)\right]^2}{VAR(H)} \quad , \qquad (2)$$

with:

$COV(\xi_k,\xi_j)$ is the covariance between ξ_k and ξ_j, n is the number of factors, h is the number of items.

Formula 2 shows that the reliability depends on the loadings of items in the scale on factor k, and on the loadings on other factors with which factor k covaries.

Model evaluation

Several criteria can be used to evaluate a factor model. Firstly, a parsimonious model is preferable to a model with more parameters. Secondly, a model should be interpretable and parameter estimates should be acceptable. For instance, the occurrence of improper solutions such as negative error variances or correlations larger than one can be indicative of a misspecified model (Van Driel, 1978; Jöreskog & Sörbom, 1989 p. 41). Thirdly, a model must account for the observed correlations/covariances. Several goodness-of-fit indices have been proposed to assess the fit of a model. Fit indices are affected by sample size (Marsh, Balla, & McDonald, 1988). Other factors such as, number of items, number of factors, influence at least some of the fit indices (Anderson & Gerbing, 1984). Fit indices do not have a clear interpretation like proportion of variance explained. The dependence on characteristics of the study and the lack of a clear interpretation, make it difficult to establish a standard of what constitutes an acceptable fit. Fit indices were therefore only used to facilitate the comparison of different models within this study.

The chi-square test statistic is often used as a fit index because it offers a statistical test for the validity of the model. With polychoric correlations as input, an accurate chi-square can only be obtained with weighted least squares estimation (Jöreskog & Sörbom, 1989 p. 193). However, the computation of the weight matrix needed in a weighted least squares estimation procedure, is not feasable with the present number of variables (h=85). This requires too much computer time (Jöreskog & Sörbom, 1988 p. 1-28) and a

sample size larger than the one in the present study (Jöreskog & Sörbom, 1988 p. 3-32). Unweighted least square estimation was used instead.

An alternative approach to perform a chi-square test is to treat CBCL items as continuous variables and to compute covariances. With covariances as input, a chi-square test can also be performed with other estimation procedures. Chi-square tests performed with these estimation procedures are sensitive to departures from normality (Muthén & Kaplan, 1985). Especially in non-clinical, samples CBCL items are not normally distributed. Therefore, also with this alternative procedure no accurate chi-square can be obtained.

In the present study the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI) and the root of the mean squared residuals (RMR) were reported (Jöreskog and Sörbom, 1988, p. 44). The GFI and the AGFI are based on a comparison of the observed correlations with the correlations predicted by the factor model. The GFI and the AGFI range from 0 to 1. Larger values imply a better fit. By estimating more parameters the fit of a model can be improved. the AGFI adjusts the GFI for this statistical phenomenon. The RMR can be interpreted as the mean difference between observed correlations versus correlations predicted by the model.

Fit indices of three models were reported. A baseline model, the initial model, and the model which resulted from respecifying the initial model. The baseline model assumed that no common factors underlie the items and correlations between the items are therefore zero. Fit indices for the baseline model were used to get an impression of the lower bounds of the fit indices. The adequacy of MI's and EPC's to detect model misspecifications has been questioned (McCallum, 1986; Silvia, & MacCallum, 1988). Moreover, respecifications quided by these statistics are exploratory. Fit indices may spuriously improve because of "capitalization on chance". Both the initial model and the respecified model were therefore fitted to a validation sample. The difference in fit of the respecified model versus the initial model in the validation sample, was used to evaluate the validity of the respecifications.

Sample

Analyses were performed on CBCL data from a sample of 2,148 international adoptees living in the Netherlands. The mean age in this sample was 12.36 (standard deviation 1.17), 48.4% were boys (for a full description of this sample see Verhulst et al. 1990a,b,c). Results from the analyses were cross-validated on CBCL data from a clinical sample. This sample consisted of 1,387 children who were referred to mental health agencies in the Netherlands (see Verhulst, Akkerhuis, & Althaus, 1985, for detailed sample description). Children in the clinical sample were younger (mean age 9.75, standard deviation 3.25) compared to the children in the adoption sample, and the ratio of boys and girls was somewhat different (66.3% boys).

Results

In some cases polychoric correlations could not be computed. This problem was solved by excluding items on which more than 97% of the subjects obtained a zero score. These were the items 56a(aches,pains), 56d(eye problems), 56g(vomiting) for Somatic Complaints; 40(hears things), 66(repeats acts), 70(sees things), 85(strange ideas) for Thought Problems; 72(sets fires), 101(truancy), 105(alcohol/drugs) for Delinquent Behavior, and 97(threatens) for Aggressive Behavior.

For the 74 items that were left the means of the standardized univariate skew-

nesses and kurtoses (compared to the kurtoses of a standard normal distributed variable) were respectively 2.55 and 7.70. This shows that even with covariances as input no accurate 'chi-square' could have been obtained (Muthén & Kaplan, 1985).

In the phase of model respecifications, maximum likelihood estimation was used to obtain MI's and EPC's. In contrast to unweighted least squares estimation, maximum likelihood estimation requires an input matrix which is positive definite. The matrix with polychoric correlations failed to be positive definite, LISREL 7 therefore automatically changed the input matrix by adding a value of .1 to all elements on the main diagonal.

After excluding items on which more than 97% of the subjects obtained a zero score from the analyses, 3 items were left for Thought Problems. For these items the MI's and EPC's indicated that loadings on other constructs should be estimated too. This eventually led to underidentification of parameters associated with Thought Problems. The construct and the items which were not included in scales of other constructs by Achenbach (1991a), 9(can't get mind off thoughts), 84(strange behavior), were excluded from further analyses.

Table 1 presents fit indices of the baseline model, initial model, and respecified model. The models were fitted to the correlations between the 72 items that were left.

Tabel 1. Model fit indices for the cross-informant syndrome constructs.

		adoption sample			*clinical sample*		
Model	df.	GFI.	AGFI.	RMR.	GFI.	AGFI.	RMR.
Baseline	2556	.090	.064	.373	.153	.129	.275
Initial	2458	.956	.953	.082	.892	.884	.098
Respecified	2355	.984	.982	.050	.948	.942	.068

Note. An unweighted least squares estimation procedure with polychoric correlations as input, was used. Number of items is 72. Df. denotes degrees of freedom. Size adoption sample is 2148, size clinical sample is 1387.

The initial model was a clear improvement over the baseline model. All fit indices indicated that the final model offered the best description of the test structure. The AGFI suggested that this better fit was not only because more parameters were estimated in this model. The higher values of the fit indices of the final model in the clinical sample supported the validity of the respecifications on the initial model. For the initial and final model fit indices were lower in the clinical sample, this may reflect the smaller size of the clinical sample (Marsh, Balla, & McDonald, 1988).

Table 2 displays factor loadings in the adoption sample, obtained from fitting the respecified model. The respecified model allowed 54 correlated errors of measurement. The loading from item 23 on Delinquent Behavior was larger than 1. It should have been between -1 and 1, because a correlation matrix was used as input. When the final model was fitted to the data of the clinical sample the loading from item 23 on Delinquent behavior did not exceed 1. This suggested that the too large loading in the adoption sample was caused by sample fluctuations. Parameter estimates were acceptable in the clinical sample too, except for one loading of item 112 which exceeded 1. Results were similar for both samples. For instance, only two factor loadings changed from a small value to a small value of opposite sign, and only 4 of the 54 correlated errors of measurement changed sign. The correlation between the factor loadings in both sample was

Table 2. Factor loadings for CBCL items in adoption sample.

	item	With.	Som.	An/Dep.	Soc.	Att.	Del.	Agg.
	Withdrawn							
42.	Likes to be alone	.623	-	-	-	-.379	-	-
65.	Refuses to talk	.774	-	-	-	-	-	-
69.	Secretive	.817	-	-	-	-.203	-	-
75.	Shy/timid	.809	-	-	-	-	-.517	-
80.	Stares blankly(AT)	.813	-	-	-	-	-	-
88.	Sulks	.283	-	.197	-	-	-	.252
102.	Underactive	.565	-	-	-	-	-	-
103.	Unhappy/sad/dep.(AD)	-	-	.858	-	-	-	-
111.	Withdrawn	.823	-	-	-	-	-	-
	Somatic Complaints							
51.	Dizziness	-	.707	-	-	-	-	-
54.	Overtired	-	.300	.400	-	-	-	-
56B.	Headaches	-	.668	-	-	-	-	-
56C.	Nausea	-	.829	-	-	-	-	-
56E.	Rashes/skin problems	-	.319	-	-	-	-	-
56F.	Stomaches	-	.874	-	-	-	-	-
	Anxious/Depressed							
12.	Lonely	-.219	-	.901	-	-	-	-
14.	Cries a lot	-	-	.628	-	-	-	-
31.	Fears impulses	-	-	.637	-	-	-	-.214
32.	Needs to be perfect	-	-	.985	-.640	-.176	-	-
33.	Feels unloved	-	-	.815	-	-	-	-
34.	Feels persecuted	-	-	-	.401	-	-	.261
35.	Feels worthless	-	-	.867	-	.207	-	-.213
45.	Nervous/tense(AT)	-	-	.567	-	.220	-	-
50.	Fearful/anxious	-	-	.575	-	-	-	-
52.	Feels too guilty	-	-	.988	-	-	-.479	-
71.	Self-conscious	.352	-	.680	-.585	-	-	-
89.	Suspicous	.432	-	-	-	-	-	.487
112.	Worries	-	.128	.937	-.384	-	-	-
	Social Problems							
1.	Acts too young(AT)	-	-	-	.256	.407	-	-
11.	Too dependent	-	-	.440	.221	.521	-.497	-
25.	Doesn't get along w. peers	-	-	-	.931	-	-	-
38.	Gets teased	-	-	-	.572	.209	-	-
48.	Not liked by peers	-	-	-	.859	-	-	-
62.	Clumsy(AT)	-	-	-	.282	.484	-	-
64.	Prefers younger kids	-	-	-	.368	.241	-	-
55.	Overweight	-	-	-	.293	-	-	-
	Attention Problems							
8.	Can't concentrate	-	-	-	-	.887	-	-
10.	Can't sit still	-.271	-	-	-	.507	-	.337
13.	Confused	-	-	.390	-	.565	-	-
17.	Daydreams	.633	-	-	-	-	-	-
41.	Impulsive	-	-	-	-	.468	-	.472
46.	Twitches	.161	-	-	-	.322	-	-
61.	Poor school work	-	-	-	-	.355	.381	-
	Delinquent Behavior							
26.	Lacks guilt	-	-	-	-	-	.825	-
39.	Bad companions	-	-	-	-	-	.646	-
43.	Lies	-	-	-	-	-	.846	-
63.	Prefers older kids	-	-	-	-	-	.205	.242
67.	Runs away from home	-	-	-	-	-	.782	-
81.	Steals at home	-	-	-	-	-	.752	-
82.	Steals outside home	-	-	-	-	-	.767	-
90.	Swearing/obscenity	-	-	-	-	-	-	.795
96.	Thinks too much ab. sex	-	-	-	-	-	.239	.420
106.	Vandalism	-	-	-	-	-	.630	.270
	Aggressive Behavior							
3.	Argues	-	-	-	-	-	-	.735
7.	Brags	-	-	-	-	-	-	.673

(continued)

Table 2. (continued)

	item	With.	Som.	An/Dep.	Soc.	Att.	Del.	Agg.
16.	Bullies	-	-	-	-	-	.563	.285
19.	Demands attention	-	-	-	-	.369	-	.557
20.	Destroys own things	-	-	-	-	-	.827	-
21.	Destroys others' things		-	-	-	-	.640	.249
22.	Disobedient at home	-	-	-	-	-	.454	.427
23.	Disobedient school	-.301	-	-	-	-	1.017	-
27.	Jealous	-	-	.281	-	-	-	.347
37.	Fights	-	-	-	-	-	.184	.627
57.	Attacks people	-	-	-	-	-	.342	.598
68.	Screams	-	-	-	-	-	-	.776
74.	Shows off	-	-	-	-	.374	-	.412
86.	Stubborn/irritable	.473	-	-	-	-.289	-	.652
87.	Sudden mood changes	-	-	.369	-	-	-	.388
93.	Talks too much	-.691	-	.509	-	.382	-	.282
94.	Teases	-	-	-	-	-	-	.778
95.	Temper tantrums	-	-	-	-	-	-	.757
104.	Loud	-.259	-	-	-	.428	-.368	.929
	reliability	.82	.89	.87	.81	.81	.93	.93
	Cronbach's alpha	.75	.55	.80	.78	.72	.86	.84

Note. With.=Withdrawn, Som.=Somatic Complaints, An/Dep.=Anxious/Depressed, Soc.=Social Problems, Att.=Attention Problems, Del.=Delinquent Behavior, Agg.=Aggressive Behavior. Items are listed in the scales as reported by Achenbach (1991a, pp. 48-51). Factor loadings that are underlined denote items selected for that scale using the criteria mentioned in the text. Reliability is computed from formula 2 in combination with polychoric correlations. Cronbach's alpha was based on the covariances. An unweighted least squares estimation procedure was used, with polychoric correlations as input. Sample size is 2148. A factor loading fixed at zero is denoted by -.

.93. This supported the validity of the respecifications. Factor loadings were smaller in the clinical sample. The mean absolute factor loading was in the clinical sample .44, in the adoption sample .51.

Table 3 shows the estimated correlations between the 7 factors in the adoption and clinical sample. No correlation was close to one, and each construct showed a somewhat different pattern of correlations with the other constructs. This supported the validity of the cross-informant syndrome constructs, and showed that the constructs can be viewed as separate dimensions of problem behavior. Correlations were lower in the clinical sample

Tabel 3. Correlations between the cross-informant syndrome constructs.

	1.	2.	3.	4.	5.	6.	7.
1. Withdrawn	1	.227	.647	.578	.395	.429	.171
2. Somatic Comp.	.282	1	.421	.165	-.053	-.027	.053
3. Anxious/Depressed	.785	.444	1	.764	.290	.342	.439
4. Social Prob.	.674	.232	.816	1	.480	.653	.623
5. Attention Prob.	.455	.212	.546	.614	1	.554	.428
6. Delinquent Beh.	.676	.214	.721	.813	.622	1	.744
7. Aggressive Beh.	.395	.249	.669	.674	.553	.809	1

Note. An unweigthed least squares esimation procedure, with polychoric correlations, was used. Below diagonal adoption sample (N=2148), above diagonal clinical sample (N=1387).

than in the adoption sample.

Scales were constructed to evaluate the content validity of the items. An item was included in a scale of a construct when it met three criteria. First, to select an item for the scale of a construct, the factor loading had to be larger than .4. An Item with a loading above .4 was considered to be a valid indicator of that construct. Second, an item was not allowed to have loadings on other constructs larger than .4 or less than -.4. An item with loadings larger than .4 or smaller than -.4 on other constructs as well, was excluded because it was considered to be a valid indicator of more than one construct. As a result of this criterion, items cannot be included in more than one scale, and non-specific indicators are excluded. It is not desirable that an item appears in the scales of more than one construct; it would, for instance, spuriously inflate the observed correlation(s) between the constructs. Finally, an item had to improve the reliability of the scale. This was determined by computing the reliability of the scale with and without the item (using formula 2). The factor loadings of items, which met these criteria, are shown in bolface and are underlined in Table 2.

Table 2 shows differences between scales reported by Achenbach (1990a), and scales derived for the adoption sample. In the present study, 26 items were not selected for a scale of one of the cross-informant syndrome constructs. These items were according to our criteria not indicators of just one construct, or did not improve the reliability of the scale. Six items were selected for another scale compared to the scales reported by Achenbach (1990a).

Polychoric correlations were used to compute the reliability using formula 2. For sake of comparison Cronbach's alpha was also reported. Alpha was based on product moment correlation coefficients (pmcc), as is usually done. Table 2 shows that alpha was lower in all cases. When alpha was computed with polychoric correlations, there was no systematic difference between alpha and formula 2. This indicated that the lower alpha's in Table 2 were caused by the use of pmcc's instead of polychoric correlations. With ordinal variables the use of pmcc's leads to underestimates of the reliability.

Cronbach's alpha's based on pmcc's were always higher than reliability estimates computed with formula 2 using pmcc's. Alpha was less affected by the attenuation of the correlations, caused by the use of ordinal instead of continuous response scales.

Discussion

Confirmatory factor analysis was used to study the validity of Achenbach's cross-informant syndrome constructs in a sample of international adoptees. Factor loadings were used to study the content validity of the items.

Fitting a model based on the cross-informant syndrome constructs yielded interpretable results and acceptable parameter estimates in the sample of international adoptees. These results supported the validity of 7 of the 8 constructs that were studied.

The content validity of the items was evaluated by constructing scales for the cross-informant syndrome constructs. Differences were found between scales derived in this study and those reported by Achenbach (1991a). Some of these differences may be attributed to the use of different item selection procedures. For instance, in his analyses Achenbach (1991a) allowed items to be selected for more than one scale. Other differences may reflect specific properties of the samples. For instance, in the present study eleven items were excluded because polychoric correlation could not be computed for these items. All these items had very low variances (more than 97% of the subjects

obtained a zero score). Items with low variances contain little information, because they fail to discriminate between individuals. The variance of an item is therefore also a property that should be considered in item selection (Crocker, & Algina, 1986, p. 311). Achenbach (1991b, p. 35) also excluded items from his analyses because they failed to meet a variance criterion. His analyses were performed on a clinical sample. Children in clinical sample obtain higher scores. This explains why items which meet a variance criterion in the clinical sample, failed to meet a variance criterion in the adoption sample.

The scale of Thought Problems did not seem to be suitable for studying this construct in the adoption sample. Four items had low variances. The three items with sufficient variance, were not specific to Thought Problems and loaded on other constructs too.

The factor structures of the clinical and the community sample were alike. However, compared to the adoption sample, factor intercorrelations and factor loadings were lower and error variances were higher in the clinical sample. This could reflect the 'restriction of range'. Because of the lower factor loadings and factor intercorrelations, reliabilities will be lower in the clinical sample. For example, if a scale was constructed for Withdrawn using the same criteria as for the adoption sample, the scale comprised the same items. Reliability, computed using formula 2 with polychoric correlations, equaled .74 in the clinical sample versus .82 in the adoption sample. Cronbach's alpha based on correlations equaled .61 in the clinical sample versus .75 in the adoption sample.

In conclusion, the present study supported the validity of the cross-informant syndrome constructs in a sample of international adoptees. The contribution of a number of items to the scales of the cross-informant syndrome constructs was questionable. These items had very low variances, were not indicators of just one construct, or did not improve the reliability of the scale.

4

A Study of Problem Behaviors in 10- to 15-Year-Old Biologically Related and Unrelated International Adoptees[1]

Edwin J.C.G. Van Den Oord[2,3], Dorret I. Boomsma[3], and Frank C. Verhulst[2].

Abstract

In the present paper, genetic and environmental influences on problem behaviors were studied in a sample of international adoptees. Parental ratings of childrens' problem behaviors were obtained with the Child Behavior Checklist (CBCL). The sample (mean age 12.43 years) comprised a group of biological siblings (111 pairs), a group of non-biological siblings (221 pairs), and a group of singletons (94). Non-shared environmental influences were the most important source of variations in the problem behaviors. Genetic influences were substantial for externalizing behaviors, but unimportant for internalizing behaviors. Shared environmental influences accounted on average for 18% of the variance. For the CBCL total problem score, Attention Problems, and externalizing behaviors results from the present study were in agreement with findings from twin studies. The lack of genetic influences on internalizing behaviors was in contrast with results from twin studies. For the Externalizing grouping, Delinquent Behavior, and Aggressive Behavior, variances for singletons were significantly smaller than for siblings. Model fit indices indicated that these differences in variances are better attributed to smaller effects of factors associated with sibship size, than to active influences of siblings on each other. Significant sex differences were found for 7 of the 10 scales. The larger variances for boys on the Externalizing grouping and Aggressive Behavior were caused by genetic influences.

Introduction

Both in psychiatry and genetics there is an increasing interest in the study of genetic factors underlying child psychiatric conditions (Plomin, in press; Rutter, 1991; Rutter et al., 1990a, 1990b). The study of genetic factors requires special samples. Samples that provide the opportunity to separate genetic and environmental influences may be atypical in important ways (Rutter et al. 1990a; Rutter, & Redshaw, 1991), or suffer from systematic biases. The generalization of findings to the general population may therefore be limited, or the conclusions concerning genetic influences biased. The implication of this is that multiple methods should be employed. Although all strategies

---

[1] The authors are grateful to Mrs. Herma Versluis-den Bieman for her helpful comments

[2] Department of Child and Adolescent Psychiatry, Sophia Children's Hospital-Erasmus University Rotterdam, the Netherlands

[3] Department of Psychonomics, Free University, Amsterdam, the Netherlands

suffer from limitations, they do not necessarily have the same ones (Rutter et al. 1990a).

Most problem behaviors in children and adolescents do not form clearcut diagnostic categories. Problem behaviors in children generally involve quantitative variations of behavior that most children display to some degree. It is therefore preferable to examine genetic influences on problem behaviors assessed as quantitative variations of behavior rather than all-or-none categories. From a genetic point of view it is likely that for these continuous variations the effects of many genes are involved (McGuffin, & Gottesman, 1985), and that methods of quantitative genetic theory have to be applied for studying child psychiatric conditions.

In the classical genetic design, sibling resemblance is viewed as caused by the 'passive' sharing of genes and environments. A number of authors have suggested that this passive view may be too simplistic (Carey, 1986; Dunn, 1983; Eaves, 1976; Patterson, 1982). For instance, by imitating each other's behaviors, siblings may become more alike. The probable importance of these kind of active influences from one sibling on the other have been noted in the area of juvenile delinquency (Rowe, 1983; Shields, 1977), and evidence for such influences on adult delinquency (Carey, 1992) and boys' externalizing behaviors has recently been found (Neale, & Cardon, 1992, p. 205).

Rutter (1970, pp. 222-223) found significant associations between sibship size and several problem behaviors. One possible explanation would be that active influences from siblings on each other are important for other problem behaviors too. However, sibship size could also be a beneficial or harmful factor by itself. For instance, it seems reasonable to suppose that as the number of children in the family increases there would be a decrease in the amount of time parents spend with any child (e.g. Patterson, 1982 p. 22). On the other side, children appear to benefit from both offering and receiving comfort from siblings (Dunn, & McGuire, 1992). In a genetically informative study design, it is possible to examine whether relations between sibship size and problem behaviors are caused by active influences from siblings on each other or if sibship size simply represents an aspect of a shared environment from which children are passive recipients.

The sample in the present study comprised 3 groups of international adoptees living in The Netherlands: 2 groups of sibling pairs and 1 group of singletons. The first group of siblings consisted of adoptees who were biologically related, the second group of siblings consisted of biologically unrelated adoptees. These groups enabled us to study genetic and environmental influences on problem behaviors. The group of adopted biological siblings is rather unique. In most sibling adoption designs, the difference in resemblance of adopted children and the biological children of adoptive parents versus the resemblance of the non-adopted biological siblings is used to study genetic influences. Thus, adopted children are usually compared with controls who are raised by their biological parents, while in our study both groups are raised by adoptive parents. The final group consisted of adoptees who grew up as singletons. Such a group of singletons is unique to adoption samples, and can be used to study the influence of multiple children within one family on problem behaviors.

Twin data were used in the majority of genetic studies of problem behaviors in children. For a number of behaviors such as anxiety, depression, and aggression no adoption study has even been reported yet. Although biases are probably present in the adoption sample used in the present study, these biases may be quite different from possible biases in twin samples (e.g. an exaggaration of MZ twin resemblance, assortative mating). The adoption sample in the present study therefore provided an opportunity for a comparison with twin study inferences about genetic and environmental effects on

problem behaviors in children. If the same results are obtained, conclusions are more likely to be valid.

Method

Assessment instrument

Parental ratings of children's problem behaviors were obtained with the CBCL/4-16 (Child Behavior Checklist for children aged 4 to 16). The CBCL consists of 120 items, which describe a broad range of problems of concern to parents and clinicians. Parents are requested to circle a 0 if the problem is not true of a child, a 1 if the item is somewhat or sometimes true and a 2 if it is very true or often true.

Achenbach (1991) derived so-called cross-informant syndrome constructs. These constructs each describe a relatively narrow range of problem behavior and were labeled: Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior. The cross-informant syndrome constructs are similar for different types of informants, both sexes, and age groups 4-11 and 12-18. Thought Problems was not studied in the present paper, because frequencies of problems comprising this syndrome were too low.

The applicability of the cross-informant syndrome constructs in the sample of international adoptees was studied in an earlier paper (Van Den Oord, Verhulst, & Boomsma, submitted). Results supported the validity of the constructs but suggested some modifications of the scales. These modified scales were used in the present study.

In addition to the cross-informant syndrome constructs, the total problem score and the Internalizing/Externalizing groupings of problem behavior were studied. The total problem score is the sum of all 118 close ended items. The total problem score contains items not present in the scales for one of the cross-informant syndrome constructs. The Internalizing/Externalizing groupings were included in the present study because similar groupings of problem behaviors appear frequently in child clinical literature (Achenbach, 1991 p. 63). The Internalizing score was obtained by summing the scores of the Withdrawn, Somatic Complaints and Anxious/Depressed scales. The Externalizing score was obtained by summing the scores of the Delinquent Behavior and Aggressive Behavior scales.

Sample

The sample was part of a larger sample of 2,148 international adoptees living in the Netherlands (for a full description of this sample see Verhulst, Althaus, & Versluis-den Bieman, 1990a,b). This subset consisted of 758 adoptees (mean age is 12.43, SD = 1.16), of whom 385 were girls and 373 were boys. The mean age at placement in Dutch adoptive homes was 26.93 months (SD = 22.98). The sample was divided into two groups of siblings and one group of singletons. The first group of siblings consisted of 111 pairs of biologically related adoptees. This group was further divided into 35 pairs of girls, 30 pairs of boys and 46 opposite-sex pairs. The second group of siblings consisted of 221 pairs of biologically unrelated adoptees. In this group there were 48 pairs of girls, 44 pairs of boys and 129 opposite-sex pairs. The third group consisted of 94 adoptees who grew up as the only child. This group included 44 girls and 50 boys. In all, there were 8 (2x3+2) groups.

Background characteristics of the groups of biological siblings, non-biological siblings, and singletons are presented in Table 1 and Table 2 (for a detailed discussion of these variables see Verhulst, Althaus, & Versluis-den Bieman, 1990a, 1990b, 1992).

Table 1 displays the countries of origin.

Table 1. Countries of origin.

	bio. sibs	non-bio sibs	singletons
	n=222	n=442	n=94
Korea	47.3%	21.5%	20.2%
Other Asia	15.3%	20.7%	24.4%
Colombia	26.6%	13.3%	8.5%
Other Non-European	8.1%	30.3%	20.3%
Europe	2.7%	14.2%	26.6%

Table 1 shows that there were differences between the three groups concerning the country of origin.

Table 2. Means and Standard deviations on background variables.

	bio. sibs	non-bio. sibs	singletons	regression coef.
	n=222	n=442	n=94	n=2148
AGE (years)	12.5 (1.18)	12.4 (1.15)	12.5 (1.16)	.09(W), .07(I)
PLACEMENT (months)	43.5 (21.4)	20.7 (20.4)	17.2 (17.6)	-.12(AT), -.14(AG), -.1(E)
SES	4.71 (1.39)	4.63 (1.42)	4.00 (1.49)	.07(D)
CARETAKING	1.76 (.63)	1.48 (.60)	1.41 (.53)	.10(W),
NEGLECT	1.75 (.82)	1.54 (.75)	1.39 (.64)	.13(T), .11(S), .11(AT), .08(D), .11(AG), .1(E)
ABUSE	1.29 (.58)	1.13 (.43)	1.04 (.19)	.13(W), .11(AD), .1(D), .13(I), .09(E)
HEALTH	1.40 (.49)	1.41 (.49)	1.42 (.50)	.08(AD), .09(I)

Note. T is Total score, W is Withdrawn, AD is anxious/Depressed, S is Social Problems, AT is Attention Problems, D is Delinquent Behavior, A is Aggressive Behavior, I is Internalizing, and E is Externalizing. Standard deviations are in parenthesis. Background variables explained in text.

Table 2 displays age (AGE measured in years), age at placement in the adoptive home (PLACEMENT measured in months), parental occupation (SES, 1=lowest occupational level, 6=highest), the number of changes in caretaking environment (CARETAKING) the child experienced before he/she was adopted, whether the child had been neglected or abused (NEGLECT and ABUSE both with categories: 1=not, 2=somewhat, 3=severe), and the child's medical condition at the time of placement (HEALTH: 1=healthy, 2=not-healthy). For CARETAKING, NEGLECT and ABUSE about 30% of the adoptive parents were not sure about their answers; their information was not used.

Group differences on background variables will only affect the genetic analyses

when the background variables are also associated with the problem behaviors that are studied. To study associations between background characteristics and problem behaviors, log-transformed scales were regressed on the background variables presented in Table 2. In these regression analyses the total sample of 2148 adoptees was used. The stepwise selection procedure of SPSS (SPSS, 1986) was used to select the most important predictors.tors. This procedure includes an additional variable in the regression equation when it significantly improves the prediction, and examines at each step the variables already in the equation for removel. Standardized regression coefficients, are presented in Table 2. Table 2 shows that biological siblings were placed in their adoptive homes later than were the non-biological siblings or the singletons. For SES, CARETAKING, NEGLECT, and ABUSE biological siblings tended to have higher scores than non-biological siblings, and non-biological siblings tended to have higher scores than singletons. However, regression coefficients indicated that influences on problem behaviors were too small, to justify incorporating the background variables in the genetic analyses and to apply corrections for group differences on these variables (the mean multiple correlation was mean .14, which corresponds with 2% explained variance).

The sibship size was 3.3 for biological siblings and 3.0 for non-biological siblings. The mean age difference was 1.4 years for biological siblings, and 1.6 years for non-biological siblings. In 75% of the cases, the non-biological sibling pairs came from the same country of origin.

Model

The model used for data analysis is presented in equation 1 for opposite-sex pairs (subscript g refers to girls, subscript b refers to boys).

$$\begin{aligned} P1 &= sP2 + h_g A_1 + c_g C_1 + e_g E_1 \\ P2 &= sP1 + h_b A_2 + c_b C_2 + e_b E_2 \ , \end{aligned} \qquad (1)$$

In formula 1, P1 and P2 represent the scores of respectively the first (girls) and second sibling (boys). A refers to the additive genetic factor, C to the shared environmental factor, and E to the non-shared environmental factor. Parameters h, c, and e, are the loadings from P on respectively A, C, and E. Parameter s is the effect from one sibling on the other, and is not allowed to depend on the sex of the child.

sibship size effects

Possible differences in variances between the groups of biological siblings, non-biological siblings, and singletons are important to study whether siblings influence each others behavior in an active way or if sibship size represents an aspect of the shared environment.

Parameter s in model 1 represents the active influence from one sibling on the other (see Carey, 1986; Eaves, 1976; Neale, & Cardon, 1992). When s is positive siblings cooperate or imitate each others behaviors. Negative values imply contrast or competition effects. Formula 2 expresses the observed variance in case of sibling interaction (e.g. Neale, and Cardon, 1992, p. 208). When A,C and E are scaled to have variances equal to one, then :

$$VAR(PI) = \frac{(h_g^2 + 2rsh_g h_b + s^2 h_b^2) + (c_g^2 + 2sc_g c_b + s^2 c_b^2) + (e_g^2 + s^2 e_b^2)}{(1-s^2)^2} \quad , \quad (2)$$

r is the genetic correlation between siblings.

Formula 2 illustrates that the variance of scales with sibling effects and genetic influences depends on the genetic correlation, r. In addition, the variance for singletons will be different from the variance for siblings. The variance for singletons equals (s=0 in formula 2): $VAR(P)=h^2+c^2+e^2$.

Variances of groups of siblings and singletons may be different because of other reasons than active influences from siblings on each other. For instance, as the number of children in the family increases there could be a decrease in the amount of time parents spend with each child. In this case, sibship size could be viewed as an aspect of the shared environment, and there would be an association between problem behavior and the number of children in the family. Children in the groups of siblings come from families of different sizes. For sibling groups, sibship size is a variable that contributes to the variance of problem scores. For singletons sibship size is not a variable, and it can therefore not contribute to the variance. Consequently, the variance for siblings will be larger than for singletons. In terms of model 1, these kind of sibship size effects could be accounted for by estimating a separate shared environmental effect in the group of singletons. The variance for siblings and singletons can then be expressed as: $VAR(P_{sibl.})=h^2+c^2_{sibl.}+e^2$, and $VAR(P_{sing.})=h^2+c^2_{sing.}+e^2$, with $c^2_{sibl.} \geq c^2_{sing.}$

The two models for explaining differences in variances between groups of siblings singletons lead to different predictions, and are therefore testable alternatives. In contrast to a model with sibling effects, the model which views sibship size as an aspect of the shared environmental does not predict different variances for biological and non-biological siblings. Furthermore, when shared environmental effects are not important this model cannot account for a difference in variance between singletons and siblings, while a model with sibling interaction still can.

sex differences

To account for sex differences, models with general scalar sex limitation, and models allowing specific scalar sex limitation were fitted (see Heath, Neale, Hewitt, Eaves, & Fulker, 1989; Neale, & Martin, 1989). Both models assume that the same genes and environments influence behavior in girls and boys. However, in a model with general scalar sex limitation parameters h, c, and e in one sex are a constant multiple of the parameters in the other sex. In a model with specific scalar sex limitation parameters h, c, and e are estimated for girls and boys separately. The first model is more parsimonious because it estimates only one additional parameter compared to a model without sex differences, while the second model estimates three additional parameters. A model with general scalar sex limitation can account for differences in variances between girls and boys, but the relative importance of genetic (the heritability), and environmental influences is constrained to be equal. In contrast, the relative importance of genetic and environmental influences may depend on the sex of the child for a model with specific scalar sex limitation.

Model selection

To select the best fitting model, 4 variations of the model in equation 1 were fitted to the data. Models with either general scalar sex limitation or specific scalar sex limitation were in one case elaborated with a parameter for sibling interaction. In the other case sibling interaction was not allowed (s=0), but an additional parameter was estimated in the groups of singletons. This parameter could account for possible smaller contributions of factors associated with sibship size. The model with the largest probability (p-value) was preferred to draw conclusions.

LISREL 7 (Jöreskog, & Sörbom, 1989) was used to obtain parameter estimates through a simultaneous analysis of the 8 groups in the sample. LISREL requires that every group has the same number of variables. However, in the groups of singletons there is only 1 variable. To use LISREL a dummy variable D with pseudo values VAR(D)=1 and COV(P1,D)=0 was specified for the groups of singletons (analogous to the way missing data can be handled in LISREL, Jöreskog, & Sörbom, 1989 p. 259). For the 8 groups there were 20 (6x3+2) observed statistics and 4 (2x2) statistics associated with the dummy variables. The degrees of freedom were adjusted for these dummy statistics (this was done by putting df=-4 on the OU line of the last group).

The implementation of models with sibling interaction, general scalar sex limitation, and specific scalar sex limitation can be achieved by approaches illustrated by Heath, Neale, Hewitt, Eaves, & Fulker (1989), Neale and Cardon (1992), and Neale, & Martin (1989).

For models with general scalar sex limitation and that allow smaller "shared environmental" effects for singletons, shared environmental effects were simply estimated for singletons separately. For models with specific scalar sex limitation and that allow smaller shared environmental effects for singletons, one additional parameter b was estimated in the groups of singletons. The shared environmental effects for girls and boys in the groups of singletons can be obtained by multiplying the shared environmental effect for girls c_g and boys c_b in the sibling groups with b. This procedure is in agreement with a model of specific scalar sex limitation, and results in the same scalar sex difference in shared environmental effects for singletons as for siblings.

CBCL syndrome scores display considerable skewness and kurtosis. To perform accurate significance tests with maximum likelihood estimation (Muthén, & Kaplan, 1985), scale scores were log-transformed.

The biological sibling were assumed to be full siblings, consequently the genetic correlation was fixed at .5.

Results

Table 3 reports variances for families of different sizes, and correlations between sibship size and scale scores. Sibship size was computed by summing all biological, adoption, and foster children in a given family.

Correlations between sibship size and scale scores were low. The absence of positive or negative correlations imply that the presence of multiple children in one family can, in general, neither be considered beneficial nor harmful. The total problem score, Delinquent Behavior, Aggressive Behavior, and the Externalizing grouping showed an substantial increase in variance up to sibship sizes of 4. This suggests that the presence of siblings results in more extreme scores.

Table 4 displays sibling correlations. For the total problem score, sibling correlations for biological and non-biological siblings were large, and somewhat larger for

biological siblings than for non-biological siblings.
For internalizing behaviors, sibling correlations were equal or even somewhat smaller for biological siblings than for non-biological siblings. For externalizing behaviors sibling correlations were clearly larger for biological siblings than for non-biological siblings.

Table 5 presents the results from various baseline models. Model 1 constrains for same-sex groups the variance of the first sibling to the variance of the second sibling.

Tabel 3. Variances for different sib sizes and correlations between sib size and log transformed scale scores.

sibsize	1	2	3	4	5	≥6	correlation
number of observations	n=94	n=960	n=554	n=351	n=104	n=81	n=2148
Total score	.62	.74	.79	.83	.72	.80	.00
Internalizing	.41	.39	.45	.43	.46	.35	.05
Withdrawn	.41	.44	.52	.51	.59	.39	.07
Somatic Com.	53.	50.	48.	48.	53.	47.	-.01
Anxious/Depr.	.50	.49	.52	.57	.52	.44	.01
Social Problems	86.	87.	94.	109	116	80.	.03
Attention Problems	.10	.12	.11	.12	.13	.10	-.01
Externalizing	.45	.51	.60	.66	.63	.60	.00
Delinquent Beh.	.79	.86	1.0	1.1	1.2	.96	.06
Aggressive Beh.	.23	.25	.28	.32	.31	.28	-.03

Tabel 4. Observed correlations for log transformed scores.

	biological sibs			*non-bio. sibs*		
	girls	boys	girls/boys	girls	boys	girls/boys
	pairs=35	p=30	p=46	p=48	p=44	p=129
Total score	.590	.519	.638	.566	.475	.339
Internalizing	.156	.152	.312	.414	.441	.280
Withdrawn	.139	.152	.064	.310	.130	.127
Somatic Comp.	.260	.254	-.006	.538	-.118	.080
Anxious/Depr.	.080	.213	.328	.199	.327	.229
Social Problems	.280	.141	.294	.234	.347	.117
Attention Problems	.143	.169	.465	-.126	.089	.086
Externalizing	.425	.463	.516	.372	.190	.114
Delinquent B.	.148	.418	.452	.304	.266	.123
Aggressive B.	.446	.404	.384	.211	.024	.046

Especially for Social Problems and Delinquent Behavior the fit of model 1 was poor. This poor fit was probably caused by chance, because a random procedure was used to determine the first and second sibling. Model 2 constrains variances to be equal for biological siblings, non-biological siblings, and singletons. Model 2 is nested within model 1, the chi-square difference test could was used to test model 1 against model 2 (p=.20). The decrease in fit was significant for the Externalizing grouping, Delinquent Behavior, and Aggressive Behavior. For these scales, variances were not equal across groups.

Tabel 5. Chi-squares obtained from fitting various models.

	model 1.	model 2.	model 3.	model 4.	model 5.	model 6.
	df=4	df=12	df=13	df=13	df=15	df=15
Total problem score						
	1.76	8.63	9.40	8.84	11.11	11.61
	(.780)	(.735)	(.742)	(.785)	(.745)	(.708)
Internalizing grouping						
	2.95	8.46	12.45	12.45	12.50	12.50
	(.566)	(.748)	(.491)	(.491)	(.641)	(.641)
Withdrawn						
	6.58	15.58	16.52	16.30	17.50	17.00
	(.160)	(.211)	(.222)	(.233)	(.290)	(3.19)
Somatic Complaints						
	4.38	13.39	18.44	17.11	28.26	28.26
	(.357)	(.341)	(.142)	(.194)	(.020)	(.020)
Anxious/Depressed						
	3.42	9.11	9.75	9.62	10.22	10.20
	(.490)	(.693)	(.715)	(.724)	(.806)	(.807)
Social Problems						
	9.73	20.56	21.37	22.06	22.57	22.24
	(.045)	(.027)	(.066)	(.054)	(.094)	(.102)
Attention Problems						
	1.03	9.75	13.97	13.68	14.13	13.81
	(.905)	(.638)	(.376)	(.397)	(.516)	(.540)
Externalizing grouping						
	3.96	19.73	20.41	16.19	23.45	21.35
	(.412)	(.072)	(.085)	(.239)	(.075)	(.126)
Delinquent Behavior						
	10.08	24.43	24.42	23.91	25.33	24.47
	(.039	(.018)	(.027)	(.032)	(.046)	(.057)
Aggressive Behavior						
	.80	14.08	13.94	10.52	16.97	15.85
	(.938)	(.296)	(.378)	(.651)	(.321)	(.392)

Note. Model 1. constrains the variances for the first and second sibling to each other, for same-sex sibling groups. Model 2. equals variances across groups, for girls and boys separately. Model 3. allows specific scalar sex limitation, and sibling interaction. Model 4. allows specific scalar sex limitation, an additional parameter in the groups of singletons, but no sibling interaction. Number that is underlined denotes selected model. Model 5. allows general scalar sex limitation, and sibling interaction. Model 6. allows general scalar sex limitation, an additional parameter in the groups of singletons, but no sibling interaction. Number in boldface denotes preferred model. Probabilities are in parenthesis, df. denotes degrees of freedom.

Models 3 and 4 are models with general scalar sex limitation. Model 3 allows sibling interaction. Model 4 does not allow sibling interaction, but estimates different shared environmental effects for siblings and singletons. Models 5 and 6 are models with specific scalar sex limitation. Model 5 allows sibling interaction. Model 6 does not allow sibling interaction, but estimates parameter b to account for smaller shared environmental

effects in the groups of singletons. Parameter estimates obtained from fitting model 3, 4, 5, and 6 are shown in appendix 3.

Table 5 shows that group differences in variances for the Externalizing grouping, Delinquent Behavior, and Aggressive Behavior siblings were more likely to be caused by smaller shared environmental effects in the groups of singletons, than by active influences

Tabel 6. Parameter estimates from fitting the preferred model.

	General scalar sex limitation						
	df.	χ^2	p.	h	c	e	s	scalar
Internalizing grouping								
	15	12.50	.641	.000	.347/.336	.532	-	1.048
	18	13.38	.769	-	.354	.544	-	*1.000*
Withdrawn								
	15	17.00	.319	.000	.250/.096	.616	-	1.056
	18	18.75	.407	-	.253	.629	-	*1.000*
Anxious/depressed								
	15	10.20	.807	.134	.333/.313	.589	-	1.093
	17	10.25	.893	-	.335	.601	-	1.085
Social Problems								
	15	22.24	.102	3.60	3.87/2.55	7.31	-	1.138
	17	23.22	.142	-	4.08	7.99	-	1.136
Attention Problems								
	15	13.81	.540	.217	.079/.000	.217	-	1.104
	17	14.78	.611	.242	-	.205	-	1.105
Delinquent Behavior								
	15	24.47	.057	.527	.363/.042	.550	-	1.257
	16	24.47	.080	.528	.362/ -	.549	-	1.257

	Specific scalar sex limitation									
	df.	χ^2	p.	h_g	h_b	c_g	c_b	e_g	e_b	s	b
Total problem score											
	13	8.84	.785	.385	.592	.639	.487	.423	.443	-	.671
	17	14.42	.637	*.479*	*.479*	*.549*	*.549*	*.448*	*.448*	-	*1.00*
Somatic Complaints											
	13	17.11	.194	-.250	4.36	4.42	.524	5.61	5.03	-	1.37
	17	19.62	.294	-	3.60	4.63	-	*5.62*	*5.62*	-	*1.00*
Externalizing grouping											
	13	16.19	.239	.444	.742	.420	.173	.263	.197	-	.000
	15	16.22	.368	.453	.725	.419	.176	*.250*	*.250*	-	-
Aggressive Behavior											
	13	10.52	.651	.319	.499	.256	.040	.183	.242	-	.000
	16	10.88	.817	.322	.526	.253	-	*.180*	*.180*	-	-

Note. Parameter h is additive genetic effect, c is shared environmental effect, e is non-shared environmental effect, s is sibling effect. Subscript g refers to girls and subscript b refers to boys. Number in italics denotes estimate constrained to be equal for both sexes or fixed at that value, - denotes parameter fixed at zero. For models with scalar sex limitation: scalar girls is fixed at 1, estimate of shared environmental effect before slash is appropriate to siblings, estimate after slash is appropriate to singletons. For models with specific scalar sex limitation, parameter b is fixed at 1 for siblings and estimated for singletons.

from siblings on each other. For the Internalizing grouping, and Anxious/Depressed model 5 and model 6 fitted equally well. This occurred because there were no group differences in variances. An equal fit could therefore be obtained by estimating a near zero sibling effect (model 5), or almost equal shared environmental effects for singletons and siblings (model 6). These findings indicated that the choice which model to prefer was trivial, because there were neither sibling effects nor represented the presence of siblings an aspect of the shared environment.

To simplify the preferred model, a variety of more restrictive models were fitted to the data. In these models parameters were fixed at zero or constrained to be equal for both sex groups. The chi-square difference test was used ($p=.20$), to test whether the restrictions of the more parsimonious model were appropriate. Non-shared environmental influences include errors of measurement. Scales cannot be expected to be perfectly reliable, non-shared environmental influences were therefore never fixed at zero. Estimates of the parameters of the preferred model and the model that resulted from simplifying the preferred model are shown in Table 6.

For internalizing problems non-shared environmental influences were largest, and genetic influences were smallest. For externalizing problems genetic influences were larger than either shared or non-shared environmental influences. Sex differences were significant for 7 of the 10 scales. For the Externalizing grouping and Aggressive Behavior, genetic influences were larger and shared environmental influences were smaller for boys.

Discussion

Adoption data were used to study genetic influences on problem behaviors. Most genetic studies of problem behaviors in children have used twin data. Twin studies may suffer from limitations, not present in adoption studies. The adoption sample from the present study provided an opportunity for a comparison with twin study inferences about genetic and environmental effects on problem behaviors in children.

For the total problem score, Attention problems, and externalizing behaviors results from the present study were in agreement with findings from twin studies, thereby strengthening twin study inferences about genetic influences on these syndromes. The lack of genetic influences on internalizing behaviors was in contrast with results from twin studies.

Total problem score. A study by Edelbrock, Rende, and Plomin (1992) comprised 99 pairs MZ and 82 pairs of same-sex DZ twins (mean age 11.0 years). Ratings of twins' problem behaviors were obtained with the CBCL. Genetic, shared environmental, and non-shared environmental influences accounted for 32%, 48%, and 20% of the variance of the total problem score. In the present study these percentages were respectively 31%, 41%, and 27%, which is in close agreement with the results from the twin study by Edelbrock, Rende, and Plomin (1992). A twin study by Graham, and Stevenson (1985), and three twin studies reported by Shields (1977) also found evidence of genetic influences on a general measure of psychiatric dysfunctioning in children. However, compared to the CBCL total problem score, these other measures suggested smaller shared environmental influences.

Attention problems For Attention Problems genetic influences accounted for 47% of the variance, shared environmental influences were very small. This finding is in close agreement with the CBCL study by Edelbrock, Rende, and Plomin (1992). Furthermore, it also is in agreement with the majority of twin studies of hyperactivity/activity (Goo-

dman, & Stevenson, 1989b; O'Connor, Foch, Sherry, & Plomin, 1980; Matheny, & Dolan, 1980; Plomin, 1986 p. 214; Torgersen, 1982; Willerman, 1973).

Externalizing behaviors. For the Externalizing grouping genetic influences accounted for 65% of the variance, the remainder consisted of on the average equal parts of shared and non-shared environmental influences. The CBCL study by Edelbrock, Rende, and Plomin (1992), and a twin study of the Bullying scale from the Connors Parent Symptom Rating questionnaire (O'Connor, Foch, Sherry, & Plomin, 1980) yielded similar results. However, a CBCL study by Hewitt, Silberg, Neale, Eaves, and Erickson (1992) with 414 MZ and 569 DZ twin pairs suggested that shared environmental influences accounted for more than 60% of the variance, and genetic factors 38%.

In the present study, genetic influences were especially high for Aggressive Behavior. A higher heritability for Aggressive Behavior than for Delinquent behavior was also found in the twin study by Edelbrock, Rende, and Plomin (1992). An earlier review by Plomin 1990b showed an inconsistent pattern for results from twin studies of aggression in childhood and adolescents. Interestingly, results from the present study resembled those from the only CBCL study in Plomin's review. Ghodsian-Carpey, and Baker (1987) found that genetic influences accounted for over 90% of the variance of a CBCL Aggression scale. This agreed with the 70% found in the present study.

A heritability of .39 as was found in the present study for Delinquent Behavior, is in agreement with findings from twin studies (Edelbrock, Rende, and Plomin, 1992; Rowe, 1983). McGuffin and Gottesman (1985) pooled the findings of 6 twin studies concerning juvenile delinquency and crime. They found, for the total of 83 pairs of MZ twins and 61 DZ twins, a weighted average concordance rate of 87% for MZ twins and 72% for DZ twins. Their review also suggested a genetic component, but shared environmental influences were clearly larger in that study.

Internalizing behaviors. Genetic influences on the Internalizing grouping and scales that constitute this grouping were small or absent. This finding is in contrast with twin studies of the CBCL Internalizing grouping (Hewitt, Silberg, Neale, Eaves, &, 1992); Edelbrock, Rende, & Plomin (1992). In addition, twin studies of for example anxiety and depression also showed substantial genetic influences (Gottesman, 1963, 1965; Scarr, 1966; Stevenson, Batten, & Cherner, 1992, Wierzbicki, 1987). Compared to twin studies, genetic influences were smaller and non-shared environmental influences were larger in the present study.

Some biases may have affected the results from the present study. Firstly, it was assumed that the biological siblings were full siblings. However, within this group there could be half siblings. If the group of biological siblings consisted of half siblings rather than full siblings the genetic correlation would be .25 instead of .50. To check our assumption that the biological siblings were full siblings, analyses were also performed with a genetic correlation of .25. For scales which showed no genetic influences (e.g. internalizing behaviors) parameter estimates and fit indices were identical to findings obtained from fitting models which assumed a genetic correlation of .5. For scales which showed genetic influences, assuming a genetic correlation of .25 yielded larger genetic effects and smaller non-shared environmental effects. Estimates of the shared environmental effects were hardly affected. Models assuming a genetic correlation of .25 yielded unacceptable high heritabilities compared to the findings from twin studies such as reported above. Moreover, for some scales zero or very small non-shared environmental effects were estimated (e.g. total problem score, Attention problem, Aggressive behavior). Very small or zero non-shared environmental effects are not plausible, because

non-shared environmental influences are confounded with errors of measurement. Finally, the fit for Aggressive Behavior was poorer for the model that assumed a genetic correlation of .25 than for the model that assumed a genetic correlation og .5. The difference found between the correlations of biological and non-biological siblings for Aggressive Behavior was too large to be consistent with a model that assumed that the biological siblings were half siblings. These analyses with a genetic correlation of .25 suggested it was not likely that a large proportion of the biological siblings were half siblings.

Secondly, it was assumed that the common environments were similar for the two groups of siblings. However, this may not be true for the time before these children were adopted. The biological siblings may have experienced more equal environments than the non-biological siblings. This could have increased similarity in the former group compared to the latter. Such differences in environments, would result in overestimates of the heritabilities. Results from the present study indicated that for scales which deviated from results of twin studies, heritabilities were lower and not higher. This suggested that the bias introduced by more similar early environments for biological siblings than for non-biological siblings is not likely to be substantial.

Thirdly, reports of Verhulst, Althaus, & Versluis-den Bieman (1990a, 1990b) showed that there were some ethnic differences in problem behaviors. Ethnic differences could have raised the sibling correlations, because biological siblings and in most cases also the non-biological siblings came from the same country. Raised sibling correlations result in overestimates of the shared environment and underestimates of the non-shared environment. Compared to findings from twin studies, results from the present study did not suggest that shared environmental influences were overestimated. It is therefore not likely that ethnic differences had a large impact on the results from the present study.

Finally, heritability estimates are population dependent. Adoption samples may deviate from twin samples, and could therefore yield a different heritability. Adopted children may show an increased genetic vulnerability (Verhulst et al., 1990b; Rutter et al., 1990a), and often have experienced more negative environmental influences (discontinuous caretaking, deprivation/abuse, malnutrition and medical conditions) which may put them at elevated risk for maladjustment (Verhulst, 1992). On the other hand, the selection of 'suitable' adoptive homes may also affect the heritability found for problem behaviors in adoption samples (Tizard, 1977; Verhulst, 1991).

For the Externalizing grouping, Delinquent Behavior, and Agressive Behavior, significant differences in variances between siblings and singletons were found. Some background variables showed larger variances for siblings than for singletons. However, associations between measured background variables and problem scores were too small to account for the differences in variances that were found. Furthermore, variances tended to increase with sibship size. This suggested a systematic effect associated with the number of siblings. Model fit indices indicated that these differences in variances are better attributed to smaller effects of factors associated with sibship size, than to active influences of siblings on each other. The low correlations between sibship size and scale scores suggested that in general sibship size influences can neither be considered harmful nor beneficial. Relations between sibship size problem behaviors appeared to be more complex. For instance, it could be that the presence of multiple children may be beneficial in one situation, or family, but harmful in another. Indeed, this would predict smaller variances for singletons, but not lower levels of problem behaviors.

Sex differences, were found for most problem behaviors. Sex differences were most obvious for the Externalizing grouping and Aggressive Behavior. For boys, genetic

influences were larger, and shared environmental influences smaller. The larger genetic effects, explained the larger total variance for boys. Under assumption that quantitative test scores are liabilities or 'risks' to behavior problems, differences in variances may have implications for prevalence rates. Externalizing problems are more prevalent in boys (Verhulst, & Koot, 1992). The larger genetic effects for boys could contribute to this larger prevalence, because more it implies that more boys are at high risk for externalizing problems.

5

Dimensions of Problem Behavior Among Young Preschoolers: Factor Structure of the Child Behavior Checklist/2-3

Hans M. Koot[1], Edwin J.C.G. Van Den Oord[1,2], Frank C. Verhulst[1], Dorret I. Boomsma[2].

Abstract

The factor structure of the Child Behavior Checklist for Ages 2-3 (CBCL/2-3; Achenbach, 1992) was investigated with three different samples - children referred to mental health services, children from the general population, and a sample of twin pairs. A series of exploratory and confirmatory factor analyses indicated a seven-factor model for all three samples. Syndromes were labeled Oppositional, Withdrawn / Depressed, Aggressive, Anxious, Overactive, Sleep Problems, and Somatic Problems. Internal consistency estimates, test-retest stability, and interparent agreement were moderate to high for the seven factors. Factor intercorrelations and a second-order factor analysis provided support for two groupings of problem behaviors - Externalizing and Internalizing.

Introduction

In recent years preschoolers' problem behaviors have received considerable attention from clinicians and researchers (e.g., Campbell, 1990; Richman & Lansdown, 1988; Trad, 1988). However, research and clinical efforts are impeded by difficulties in defining criteria for deviance and by a lack of knowledge concerning the syndromes that can be distinquished among young preschoolers (ages 2 and 3 years).

For school-aged children and adolescents multivariate analyses of rating scales for assessing behavioral and emotional problems have shown that a number of dimensions of problem behaviors can be distinguished (Achenbach, 1991a; Quay, 1986). Further, broad-band groupings of problem behaviors have been identified across studies and instruments for which Internalizing and Externalizing have now become generally accepted labels (Cicchetti & Toth, 1991).

For preschoolers, only syndromes comparable to the broad-band Internalizing and Externalizing dimensions have been replicated consistently. Factor-analyses of several 3-step teacher ratings (comprising 22 to 49 items) of problem behavior in children in the age range of 2 to 6 years, have repeatedly yielded an 'externalizing' dimension labeled Hostile-Aggressive, Anger-Defiance, and Conduct-Restless-Aggressive by various authors, versus an 'internalizing' dimension labeled Anxious-Fearful, Apathy-Withdrawal, or Isolated-Immature (Behar & Stringfield, 1974; Kohn & Rosman, 1972; McGuire &

[1] Department of Child and Adolescent Psychiatry, Sophia Children's Hospital-Erasmus University Rotterdam, the Netherlands

[2] Department of Psychonomics, Free University, Amsterdam, the Netherlands

Richman, 1986). Behar and Stringfield retained a third factor comprising four items with high loadings that they labeled as Hyperactive-Distractable. However, this factor has not been replicated by other authors using the same instrument (Fowler & Park, 1979; Hoge, Meginbir, Khan, & Weatherall, 1985; Tremblay, Desmarais-Gervais, Gagnon, & Charlebois, 1987).

In contrast to earlier studies concerning the factor structure of standardized ratings of preschoolers' problem behaviors, Achenbach and coworkers (Achenbach, 1992; Achenbach, Edelbrock, & Howell, 1987) were able to further differentiate between preschoolers' behavior problems. Using parent ratings on the Child Behavior Checklist for Ages 2-3 (CBCL/2-3), Achenbach et al. performed principal component analyses in two partially overlapping samples of 2-3-year-olds in 1987 (N = 398) and in 1992 (N = 546). Both samples included children referred to mental health and special education services as well as nonreferred children. The 1992 sample, however, included children from a larger geographic area than the 1987 sample, and included only those children who had CBCL/23 scores above a certain cutpoint. In both samples six syndromes were obtained and scales were composed. The scales were labeled: Social Withdrawal (1987) or Withdrawn (1992), Depressed (1987) or Anxious/Depressed (1992). The scales Sleep Problems, Somatic Problems, Aggressive (Behavior), and Destructive (Behavior) had similar labels across the 1987 and 1992 reports. The proportion of items contained in each of the 1987 scales that reappeared in the 1992 scales for the Social Withdrawal, Anxious/Depressed, Sleep Problems, Somatic Problems, Aggressive, and Destructive scales were .36, .13, .88, .75, .47, and .57, respectively. Conversely, of the items in the 1992 scales the following proportions were also in the 1987 versions: .36, .18, 1.0, .62, 1.0, and .73. Despite these differences, r's between raw scores on 1987- and 1992-scales were .73 to .99. The moderate to high overlap among items in both versions of the Sleep Problems, Aggressive Behavior, and Destructive Behavior scales reflects that in the 1992 versions these consist largely of subsamples of items contained in the 1987 versions (r's between the '87 and '92 scores were .99, .97, and .93, respectively). The invariance of the components Withdrawn and Anxious/Depressed across samples was poorer (r's equalling .75 and .73). Although these studies showed that young preschoolers' problems as assessed by parents may provide a more differentiated picture than the rather broad internalizing and externalizing dimensions, it is important to test the factor structure of the CBCL/2-3 across different samples.

In addition, Achenbach (1992) performed second-order factor analyses on the CBCL/2-3 syndrome scales. Based on these analyses an Internalizing scale was constructed comprising the Anxious/Depressed and Withdrawn scales, and an Externalizing scale comprising the Aggressive Behavior and Delinquent Behavior scales. The loading of the Withdrawn scale, however, was of equal magnitude (0.50) on both second-order factors. Further, the observed correlation between Internalizing and Externalizing scores was high (r = .75) compared to the correlation between Internalizing and Externalizing scores for the CBCL for ages 4-18 years (mean r = .52; Achenbach, 1991b). Thus, one may either question the validity of the Internalizing/Externalizing distinction using this instrument or the correctness of the assignment of the Withdrawn scale to the Internalizing dimension. The second-order factor structure may therefore also be in need of replication.

In the present study we investigated the factor structure of the CBCL/2-3 in Dutch samples. Considering that the CBCL/2-3 may be of value as an instrument to assess psychopathology in rather diverse populations, our purpose was to compose CBCL/2-3

scales based on robust factors, i.e., based on factor solutions that are reasonably invariant across variations in the selection of subjects. To the extent that factor structures are replicable across various distinct samples of subjects, the factors have a wider range of applicability as generalized constructs. The factors would then be applicable to several populations, and could be expected to generalize to other similar populations as well. We therefore performed our analyses across three independent samples: children referred to mental health agencies; children from the general population; and pairs of monozygotic and dizygotic twins. We first report exploratory and confirmatory factor analyses of syndromes, then second-order analyses of groupings of syndromes, and compare our results with those obtained in American samples.

Method

Measures

Child Behavior Checklist for Ages 2-3 (CBCL/2-3). The CBCL/2-3 (Achenbach, 1992; Achenbach, Edelbrock, & Howell, 1987) is a 99-item instrument to obtain ratings of behavioral/emotional problems by parents or caretakers of children aged 2 and 3 years. Fifty-nine of the items have counterparts in the Child Behavior Checklist for Ages 4-18 (CBCL/4-18; Achenbach, 1991b), while the remaining items have been developed specifically for ages 2-3. The CBCL/2-3 requires fifth-grade reading skills to complete. Most respondents can complete the form in less than 10 minutes. Respondents are requested to rate the items that describe the child's behavior within the past 2 months as 2 if the item is very true or often true of the child, as 1 if the item is somewhat or sometimes true of the child, and as 0 if the item is not true of the child. On 12 open-ended items the respondent is asked to describe the behavior, making it possible to correct the scoring according to the scoring instructions when necessary, and to prevent more than one item from being scored for the same problem.

Subjects and procedures

Clinical sample. The clinical sample consisted of 426 children (284 boys, 142 girls) referred to 12 child guidance and mental health settings for behavioral and emotional problems and developmental delays. The mean age of the children was 36.1 months (SD = 8.1). Ethnicity was 79.9% Caucasian, 8.2% Surinam/Antillean, 3.1% Mediterranean countries, and 8.8% other ethnic groups. Parental educational level was coded according to a 9-step scale (Van Westerlaak, Kropman, & Collaris, 1975), which was recoded for purpose of analysis to a standard Dutch 4-step scale (CBS, 1987). The mean educational level of mothers was 2.35 (SD = 0.84), and of fathers 2.55 (SD = 0.96). Employment rate was 89% for fathers and 20% for mothers. The occupational level of parents who were employed was scored on a standard Dutch 6-step scale (Van Westerlaak, Kropman, & Collaris, 1975). The mean occupational level of mothers was 2.95 (SD = 1.44), and of fathers 3.00 (SD = 1.60). The mean maternal age was 30.3 years (SD = 5.0) and the mean paternal age was 33.9 years (SD = 5.9).

For the clinical sample, the participating settings were asked to have parents, or others in custody who came with the child, fill out the CBCL/2-3 as part of the intake procedure. In preparation of the data collection, mental health workers and office personnel who were in someway involved were instructed on the purpose and procedures of the study and on how to help parents complete the checklists. Letters of introduction, including a description of the study, informed consent forms, and CBCL/2-3s were handed over to the parents and caretakers at intake. CBCL/2-3s were filled out at the

office or at home, and checked by the mental health worker. In the clinical sample, 66.8% of the respondents were mothers, 6.2% were fathers, and 27.4% were both parents or others who were in custody of the child. Demographic information was obtained from the clinical files of the children.

Community sample. Subjects in the community sample were 420 children (215 boys, 205 girls) from a target sample of 469 children randomly selected from the population of the Dutch province of Zuid-Holland. Response rate was 91.5%. The mean age of the children was 36.4 months (SD = 7.0). Ninety-five percent of the children were Dutch, 1.6% were Surinam, 0.2% came from the Dutch Antilles, 0.2% were Turkish, and 3.1% had another nationality. The mean educational level (CBS, 1987) of mothers was 2.56 (SD = 0.80), and of fathers 2.74 (SD = 0.87). Employment rate was 93% for fathers and 32% for mothers. The mean occupational level (Van Westerlaak, Kropman, & Collaris, 1975) of mothers was 3.53 (SD = 1.41) and of fathers 3.68 (SD = 1.44). The mean maternal age was 31.5 years (SD = 4.4) and the mean paternal age was 34.2 years (SD = 5.0). Ten children (2.4%) had been referred to a child mental health agency within the past 12 months.

In the community sample, a letter was sent to the parents of the 469 eligible children explaining the purpose of the study, the way in which the child was selected, and an announcement that an interviewer would contact them. The parents were contacted by telephone, and subsequently visited by one of four trained female interviewers, who had an education at the master's level in special education or psychology. The interviewer read the CBCL/2-3 problem items regarding the target child aloud, and scored the parent's responses. In all cases the mother was the prime respondent. After completing the CBCL/2-3, the parent was asked questions about demographic characteristics of the family. The duration of the interview was 30-60 minutes.

Twin sample. Subjects in the twin sample were 1306 twin pairs (1291 boys, 1321 girls) from a target sample of 1892 3-year-old twins (73% response rate). The twins' mean age was 42.1 months (SD = 4.0). Employment rate was 98% for fathers and 29% for mothers. The mean occupational level (Van Westerlaak, Kropman, & Collaris, 1975) of mothers was 3.60 (SD = 1.37) and of fathers 3.51 (SD = 1.40). The mean maternal age was 33.0 years (SD = 3.9) and the mean paternal age was 35.6 years (SD = 4.6).

In the Netherlands, about 85% of the parents of all newborns are paid a home visit by a commercial organization which promotes certain products. During this home visit parents of twins are asked to participate in the twin register kept by the Department of Psychonomics of the Free University of Amsterdam. Forty percent of all multiple births in the Netherlands are registered. CBCLs for ages 2-3 were mailed to parents of three-year-old twins. Non-responders were sent reminders and contacted by telephone. For 73% of the twin pairs both parents filled out one CBCL/2-3 for each child. For 20% only maternal ratings were available. For 8% only paternal ratings were available. Questions about demographic characteristics were contained in the questionnaire.

Data analyses

We first performed principal factor analyses with promax rotation using the SAS (1989) statistical package. An oblique rotation was preferred because different dimensions of problem behavior tend to show positive intercorrelations. With intercorrelated factors, oblique rotations yield more easily interpretable factors (Gorsuch, 1983).

After performing principal factor analysis using unweighted least squares, we subjected the first 5 to 12 factors from the analyses to varimax and subsequently to oblique promax rotations. We examined the 5- to 12-factor rotations to identify sets of

items that consistently grouped together, i.e., factor models, along the following guidelines. First a similar factor loading pattern was sought across samples, consisting of the same set of items manifesting salient loadings on the same factor for each of the rotations. Second, items were sought which loaded highly on only one factor. Third, we avoided factors too narrow in scope, i.e. having one or two items with a very specific content. Fourth, factors that failed to replicate across solutions with a different number of factors were avoided. The items loading $\geq .30$ on the factors were listed side-by-side to identify the version of each factor that included the maximum number of high loadings which also loaded highly on the other versions. We selected the solution with the best factorial structure according to two following criteria: the solution with the highest proportion of items that consistently recurred in solutions with a different number of factors was selected; if we could not decide using the first criterion, the rotation with the highest loadings and the fewest cross-loading items was retained.

Sets of items that had loadings on corresponding factors in the three samples were used to specify a model to be evaluated in a confirmatory factor analysis (CFA). By specifying the same model in all three samples, it was possible to evaluate a factor model with the same number of factors and for which identical rotations are performed. For the CFA, the mainframe version of the computer program LISREL 7 was used (Jöreskog & Sörbom, 1989) with unweighted least squares.

Results

Two items were reported for less than 5% in all samples and were therefore excluded from the analyses: Headaches and Problems with eyes without medical cause.

To make maximum use of the available information, the mean of the parental ratings on each item was used in the twin sample. When available, missing values for one parent were substituted by the rating of the other parent.

Exploratory factor analysis of syndromes. In both the clinical and the community sample, the first seven factors found in the 7- through 9-factor solutions had nearly the same items loading $\geq .30$ on similar factors in consecutive rotations. These seven factors replicated quite well in the 9-factor solution in the twin sample. Inspection of the factor intercorrelation matrix for the oblique factors and of the loadings showed that the oblique solution was clearly preferable to the orthogonal solution for the following reasons. The correlations among factors were low to moderate in all three samples. The moderate and low loadings obtained from the oblique rotations were lower and fewer items loaded $\geq .30$ than in the varimax rotations, while the high loadings were similar in both rotations. Moreover, considerably fewer cross-loadings appeared in the oblique than in the varimax rotations, which improved the interpretability of the factors. Based on the items included in these factors we applied the following preliminary labels to the factors: Oppositional; Withdrawn/Depressed; Aggressive; Anxious; Overactive; Sleep Problems; and Somatic Problems.

Confirmatory factor analytic models of syndromes. The items included in the 7-factor solutions in the clinical and community sample and those in the comparable factors from the 9-factor solution in the twin sample were compared to select items to be included in the factor model. Items with loadings $\geq .30$ on the same syndrome in at least two of the three samples were included in the factor model to be evaluated in the CFA. Sixty-nine items were selected. For the 69 items a loading was specified for the syndrome on which they loaded $\geq .30$ in at least two of the three samples. Loadings on the other syndromes

were fixed at zero. Fourteen of the 69 items had a cross-loading of .30 to .40 in one of the samples. Of these, Clumsy had cross-loadings above .30 in two samples. All 14 cross-loadings were estimated too. Thus, a total of 83 factor loadings was estimated. Finally, all correlations between the syndromes were estimated in the model to be evaluated in the CFA.

We checked whether the pattern of factor loadings was not too restrictive, and whether correlated errors of measurement had to be specified. A respecification search similar to the one described by Jöreskog and Sörbom (1989, pp. 224-225) was followed. The modification index (MI) and expected parameter change (EPC) were used to get an indication which parameters might erroneously be fixed at zero. Then the model was fitted, and all estimated parameters were inspected. To improve the parsimony of the model, non-substantial parameters were fixed at zero again. MI's and EPC's were obtained with maximum likelihood estimation. For each item the MI's and EPC's of all fixed loadings and error of measurement correlations were inspected. When both the MI and the EPC were highest for the same factor loading or error of measurement correlation in all three samples and the EPC suggested a value larger than .10 or smaller than -.10 (standardized) in all three samples, the factor loading or error of measurement correlation was freed. If estimated factor loadings or error of measurement correlations were between -.10 and .10, they were fixed again. These respecifications yielded a model with 101 factor loadings, and 17 correlated errors of measurement.

It was not possible to perform a χ^2 test. The χ^2 test statistic is often used as a fit index because it offers a statistical test for the validity of the model. CBCL item scores display considerable skewness and kurtosis. With non-normally distributed variables, weighted least squares estimation has to be used to perform an accurate χ^2 test (Jöreskog & Sörbom, 1989). However, with the present number of variables ($n = 69$), the computation of the weight matrix needed in a weighted least squares estimation procedure would require a sample size of 7245 subjects (Jöreskog & Sörbom, 1988, pp. 3-32).

Instead of weighted least squares estimation we used unweighted least squares estimation. With unweighted least squares LISREL reports three fit indices: the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), and the root of the mean squared residuals (RMR). The GFI and the AGFI are based on a comparison of the observed correlations with the correlations predicted by the factor model. The GFI and the AGFI range from 0 to 1. Larger values imply a better fit. The AGFI incorporates a penalty function for using more parameters; it may be poorer if additional parameters result in little improvement in fit. With correlations as input, the RMR can be interpreted as the mean difference between observed correlations versus correlations predicted by the model.

Table 1 presents the fit indices of the initial model, and the model which resulted from respecifying the initial model. In addition, fit indices of a baseline model are presented. The baseline model assumes that no common factors underlie the items and that the correlations between the items are therefore zero. The baseline model was used to get an impression of the lower bounds of the fit indices.

The initial model was a large improvement over the baseline model. All fit indices indicated that the final model offered the better description of the test structure. The AGFI indicated that this improvement in fit was not merely the result of the greater number of parameters being estimated in this model. The higher values of the fit indices for the final model in all three samples may be regarded as an indication of the validity of the respecifications of the initial model. The fit indices for the initial as well as for the

Table 1. Model fit indices for CBCL/2-3 syndromes for clinical, community, and twin samples.

Model	df	GFI	AGFI	RMR
Clinical sample				
Baseline	2346	.274	.253	.194
Initial	2242	.901	.894	.072
Final	2207	.932	.926	.059
Community sample				
Baseline	2346	.380	.362	.153
Initial	2242	.917	.910	.056
Final	2207	.937	.931	.049
Twin sample				
Baseline	2346	.269	.247	.197
Initial	2242	.955	.951	.049
Final	2207	.971	.968	.039

Note. An unweighted least squares estimation procedure was used. The number of items was 69, df is degrees of freedom, GFI is Goodness of Fit Index, AGFI is Adjusted Goodness of Fit Index, RMR is Root of Mean Squared Residuals. Sample sizes were N = 426 for the clinical sample, N = 420 for the community sample, and N = 2612 for the twin sample.

final model were somewhat higher in the twin sample than in the clinical and community-samples. However, this need not indicate that the model was more appropriate for the twin sample. This difference could also reflect the larger size of the twin sample (Marsh, Balla, & McDonald, 1988).

Table 2 displays the factor loadings obtained from fitting the final model to the item correlations in the three samples. Only cross-loadings with an absolute value $\geq .30$ are shown.

In general, parameter estimates were quite acceptable in all three samples. This supported the validity of the model in the three samples. However, for the respecified model the loading of item 5. Can't concentrate on Overactive exceeded the value of 1.00 in two of the three samples. Because a correlation matrix was used as input, it should have been between 1.00 and -1.00. This improper parameter estimate might have been a consequence of sampling error or a model misspecification (Van Driel, 1978; Gerbing & Anderson, 1987; Jöreskog & Sörbom, 1989, p. 41). To avoid biased estimates of other parameters that are associated with the improper estimate, the loading of item 5 on Overactive was arbitrarily fixed at .95 in all three samples.

The results of the CFA substantiated the preliminary labels attached to the factors. Factor I was defined by high loadings of the items 81. Stubborn, 16. Demands must be met, 83. Sulks, 85. Temper tantrums, 44. Angry moods, 88. Uncooperative, 97. Whining, and 13. Cries much, reflecting oppositional and demanding behavior, and lack of emotional regulation. This factor may be labeled Oppositional. Factor II was labeled Withdrawn/Depressed because highest loadings were for items 71. Little interest, 98.

Table 2. Factor loadings for CBCL/2-3 items obtained from confirmatory factor analyses in the clinical, community, and twin sample.

	factor/items	clinical sample n=426		community sample n=420		twin sample p=1306	
I. Oppositional (17 items)							
	8. Can't wait	.397		.453		.631	
	13. Cries much	.571		.507		.590	
a	15. Defiant	.567	-.365[4]	.459		.367	.346[3]
A	16. Demands must be met	.735		.677		.786	
A	29. Easily frustrated	.551		.434		.499	
A	30. Easily jealous	.551		.433		.557	
AD	33. Feelings easily hurt	.494		.384		.553	
D	36. Gets into everything	.479		.377		.465	
A	44. Angry moods	.680		.557		.671	
A	66. Screams	.493		.382		.414	
A	69. Selfish	.306		.313		.389	
W	81. Stubborn	.817		.715		.886	
A	82. Moody	.545		.379		.550	
	83. Sulks	.641		.665		.793	
A	85. Temper tantrums	.668		.565		.704	
A	88. Uncooperative	.564		.556		.643	
AD	96. Wants attention	.515		.432	.423[5]	.461	
A	97. Whining	.588		.538		.617	
II. Withdrawn/Depressed (10 items)							
W	2. Acts too young	.583		.393		.337	
W	23. Doesn't answer	.423		.499		.567	
W	26. No fun	.574		.406		.354	
AD	43. Looks unhappy	.563		.663		.486	
	56. Clumsy	.259		.291		.169	
W	67. Unresponsive	.520		.455		.360	
W	70. Little affection	.390		.541		.285	
W	71. Little interest	.633		.631		.481	
	76. Speech problem	.127		.326		.248	
	77. Stares blankly	.456		.529		.391	
	80. Strange behavior	.384		.448		.415	
W	89. Underactive	.351		.284		.280	
AD	90. Sad	.276	.319[4]	.352		.495	
W	98. Withdrawn	.600	.367[1]	.631		.416	.364[4]
III. Aggressive (9 items)							
D	14. Cruel to animals	.483		.393		.333	
D	17. Destroys own things	.629		.502		.588	
D	18. Destroys oth. things	.631		.455		.580	
A	20. Disobedient	.383	.353[5]	.311		.397	
A	35. Fights	.666		.570		.653	
A	40. Hits	.717		.607		.661	
D	42. Hurts accidentally	.521		.534		.496	
	53. Attacks people	.728		.580		.611	
A	91. Too loud	.421		.271	.368[1]	.527	
IV. Anxious (9 items)							
	3. Afraid new things	.518		.539		.511	
W	4. Avoids eye contact	.058	.418[2]	.354		.550	
AD	10. Clings to adults	.584		.517		.586	
	21. Disturbed by change	.413		.334		.377	
	32 Fears	.442		.303		.316	
AD	37. Upset by separation	.525		.423		.613	
AD	68. Self-conscious	.477		.431		.482	
AD	73. Shy	.589		.606		.720	
AD	87. Too fearful or anx.	.590		.526		.590	
	92. Upset by new	.668		.631		.730	

(continued)

Table 2. (continued)

	factor/items	clinical sample n=426	community sample n=420	twin sample p=1306
V. Overactive (5 items)				
D	5. Can't concentrate	.950	.950 -.303[1]	.950-.329[1]
	6. Can't sit still	.834	.787	.790
	11. Seeks help	.494	.540	.606
D	59. Quickly shifts act.	.759	.668	.849
W	62. Refuses active games	.461	.472	.368
VI. Sleep Problems (7 items)				
SL	22. Doesn't w. sl. alone	.612	.452	.390
SL	38. Can't sleep	.705	.568	.572
SL	48. Nightmares	.527	.514	.592
SL	64. Resists going to bed	.710	.575	.554
SL	74. Sleeps little	.538	.529	.444
SL	84. Talks/cries in sleep	.451	.416	.467
SL	94. Wakes often	.726	.701	.622
VII. Somatic Problems (3 items)				
SO	1. Aches	.661	.578	.608
SO	12. Constipated	.421	.145	.271
SO	45. Nausea	.394	.358	.467
SO	52. Painful bowel movem.	.255	.330	.262
SO	78. Stomachaches	.484	.423	.579
SO	93. Vomiting	.246	.125	.294

Note. Factor loadings are unweighted least squares LISREL estimates. Items that were deleted from the scales are underlined. Cross-loadings are given followed by the number of the factor [in brackets] on which the cross-loading occurred. Superscripts indicate items that are comprised in the CBCL/2-3 syndrome scales constructed by Achenbach (1992): A = Aggressive; AD = Anxious/Depressed; D = Destructive; SL = Sleep Problems; SO = Somatic Problems; W = Withdrawn. N is number of observations, p is number of pairs.

Withdrawn, 43. Looks unhappy, 67. Unresponsive, 23. Doesn't answer, and 26. No fun. Aggressive behavior against people and objects characterized the items with high loadings on Factor III, such as 40. Hits, 53. Attacks people, 35. Fights, 17. Destroys own things, and 18. Destroys other's things, and thus was entitled Aggressive. Two items, 20. Disobedient, and 91. Too loud, that loaded on the Oppositional factor in the exploratory analyses migrated to the Aggressive factor in the CFA. The items loading high on Factor IV, 92. Upset by new, 73. Shy, 87. Too fearful or anxious, 10. Clings to adults, and 3. Afraid to try new things reflect anxious, fearful, and shy behavior, suggesting the label Anxious. Factor V was clearly defined by high loadings of 5. Can't concentrate, 6. Can't sit still, and 59. Quickly shifts activity, which may be adequately covered by the label Overactive. Factor VI may be entitled Sleep Problems, consisting of items that all have to do with sleep disturbances, including high loadings on items such as 94. Wakes often, 38. Can't sleep, and 64. Resists going to bed. The items with highest loadings on Factor VII concern physical complaints such as 1. Aches, and 78. Stomachaches. This factor was labeled Somatic Problems.

Results were quite similar for the three samples. Pearson correlations and RMRs (see Table 3) computed for the pairwise comparisons of factor loadings between samples as measures of the congruity of the syndromes across samples (cf. Tanaka & Huba, 1984) indicated high mean congruity for all scales except Withdrawn/Depressed. The congruity for Withdrawn/Depressed was low for all comparisons. For Anxious the congruity was

Table 3. Pearson correlation coefficients (PCC) and RMRs for pairwise comparisons of estimated factor loadings between samples.

	clinical/ community		clinical/ twin		community/ twin		mean across samples	
	PCC	RMR	PCC	RMR	PCC	RMR	PCC	RMR
Oppositional	.875	.094	.833	.084	.945	.118	.898	.099
Withdrawn/Depressed	.417	.109	-.143	.158	.237	.134	.180	.134
Aggressive	.901	.118	.818	.077	.688	.110	.820	.102
Anxious	.755	.114	.448	.173	.785	.107	.690	.131
Overactive	.971	.055	.925	.079	.892	.100	.940	.078
Sleep Problems	.787	.095	.467	.126	.947	.063	.810	.095
Somatic Problems	.999	.063	.871	.076	.847	.108	.965	.082

Note. RMR = Root of Mean Squared Residuals.

low for the clinical versus twin sample comparison. Although the mean congruity for Sleep Problems was quite acceptable, the congruity was low for the clinical versus twin sample comparison. These findings suggested that with the possible exception of Withdrawn/Depressed, the same factor structure applies to all three samples.

Reliability and Interrater Agreement

Syndrome scales were composed to be used as valid representations of the syndromes in all three samples. First, scales were constructed for each sample separately. Then, general syndrome scales were composed including the items that were selected in at least two of the three samples.

Inclusion of an item in a scale was guided by the following criteria. To select an item for the scale of a syndrome, the factor loading had to be larger than .30. Second, an item was not allowed to have loadings on more than one other syndrome larger than .30 or smaller than -.30. Finally, an item had to improve the reliability of the scale. This was determined by computing Cronbach's alpha for the scale with and without the item.

Using these criteria the following nine items had to be deleted from a scale: 15. Defiant, 56. Clumsy, 76. Speech problem, 89. Underactive, 98. Withdrawn, 32. Fears, 12. Constipated, 52. Painful bowel movements, and 93. Vomiting. One exception was made. Item 5. Can't concentrate had two cross-loadings slightly below -.30 (-.303 in the community sample and -.329 in the twin sample). However, the item was maintained in the Overactive scale because it was clearly the prime defining variable in both the exploratory and confirmatory factor analyses. Further, deleting this item resulted in a decrease of the mean reliability across the three samples from .77 to .68.

It might be argued that the composition of the syndrome scales could have been affected by the inclusion of data from both twins in the twin sample or by the use of the mean parental ratings. However, estimates of factor loadings are not affected by the interdependence of twin data. Further, replication of our analyses separately for each twin member and separately for maternal and paternal ratings yielded results that were highly comparable to those presented above[1].

Table 4 reports the reliabilities of the syndrome scales. The reliabilities were above .70 for all scales, except Somatic Problems. The low alpha for this scale is not

Table 4. Internal consistency estimates, stability coefficients, and interparent agreement for CBCL/2-3 scales.

	cronbach's alpha				interparent agreement	
	clinical sample	community sample	twin sample	test-retest coefficient	clinical/ community	twin sample
	n=426	n=420	p=1306	n=51	n=60	p=1056
Oppositional	.90	.86	.91	.88	.64	.72
Withdr./Depr.	.74	.73	.64	.60	.37	.58
Aggressive	.85	.76	.82	.85	.56	.71
Anxious	.79	.76	.83	.83	.44	.70
Overactive	.77	.77	.78	.84	.64	.69
Sleep Problems	.81	.74	.70	.76	.70	.72
Somatic Probl.	.50	.43	.59	.73	.31	.60

Note. All correlations were significant at $p < .05$ or less. N is number of observations, p is number of pairs.

surprising because it consists of only three items. Reliabilities for the Withdrawn/Depressed scale were acceptable for the clinical and community samples, but only moderate for the twin sample.

To calculate test-retest correlations, the CBCL/2-3 was completed twice over a mean interval period of 19.4 days (SD = 6.6) by 51 respondents (49 mothers, 2 fathers) randomly selected from the community sample.

To assess interparent agreement, CBCL/2-3s were independently completed by both parents for 48 cases randomly selected from the community sample, supplemented with CBCL/2-3s independently completed by both parents for 12 consecutive cases at the outpatient child psychiatry unit of the Sophia Children's Hospital. In 1056 cases of the twin sample both parents completed a CBCL/2-3 for each member of the twin pair.

Test-retest rs and parent-parent correlations are reported in Table 4. Test-retest rs of the CBCL/2-3 scores ranged from r = .88 for Oppositional to r = .60 for the Withdrawn/Depressed scale. Correlations between parents' scores were generally lower than test-retest correlations. Interparent agreement was tended to be higher in the twin sample than in the combined clinical/community sample.

Second-order groupings of syndromes

Estimated factor loadings are shown in appendix 4. The pattern of estimated factor intercorrelations was fairly similar across the three samples. The range of factor intercorrelations for the clinical sample was .121 to .703, with a mean of .363. The range of factor intercorrelations for the community sample was .069 to .712, with a mean of .341. For the twin sample the range of factor intercorrelations was .232 to .754, with a mean of .454, being somewhat higher than in the other two samples. In all three samples there were some relatively high intercorrelations among the Oppositional, Aggressive, and Overactive syndromes, and among the two syndromes Withdrawn/Depressed and Anxious. This pattern suggested the presence of two broad-band dimensions.

To study the higher-order factor structure for the seven syndrome scales, we first

applied exploratory factor analyses with varimax/promax rotations to the observed intercorrelations for each of the three samples using unweighted least squares. One- two-, and three-factor higher-order solutions were specified. In all three samples, the three-factor solution was interrupted by the program because no solution could be obtained. Both the one-factor and the two-factor solutions were replicated very well across samples.

To choose which model to prefer for representing the higher order factor structure, one- and two-factor models were specified in LISREL. Because there were only seven observed variables, weighted least squares estimation could be used to perform an accurate χ^2 test. To perform an accurate χ^2 test in the twin sample, we randomly selected one twin from each twin pair.

The one-factor model had 14 degrees of freedom. This model yielded the following fit indices: clinical sample, $\chi^2 = 133.81$ ($p = .000$), AGFI = .847; community sample, $\chi^2 = 49.81$ ($p = .000$), AGFI = .882; twin sample, $\chi^2 = 166.01$ ($p = .000$), AGFI = .820. The χ^2 tests indicated that the one-factor model had to be rejected for all three samples.

For the two-factor model, the varimax/promax rotation was implemented in LISREL. From the varimax/promax rotation one reference variable was chosen for each higher order factor (Jöreskog, 1978). The variable with the highest loading on one factor and the lowest loading on the other factor was chosen as a reference variable. In the LISREL model the loading of this reference variable was estimated for the factor on which it loaded high, and fixed at zero for the other factor. In this case Aggressive and Anxious were chosen as reference variables for the first and second factor, respectively.

The two-factor model, which had 8 degrees of freedom, yielded the following fit indices: clinical sample, $\chi^2 = 35.11$ ($p = .000$), AGFI = .930; community sample, $\chi^2 = 19.83$ ($p = .011$), AGFI = .918; twin sample, $\chi^2 = 21.37$ ($p = .006$), AGFI = .959. Fit indices indicated that the two-factor model fitted better to the data than the one-factor model. However, the χ^2 test sugggested that the fit of the two-factor model also was not entirely satisfactory. In addition, the two-factor model yielded a negative error variance for the Anxious syndrome in the twin sample. To avoid biased estimates of the other parameter the error variance of Anxious was fixed at .00 (cf. Gerbing & Anderson, 1987).

The first factor was defined by high loadings in every sample for Aggressive (mean of the standardized loadings $M = .747$), Oppositional ($M = .736$), and Overactive ($M = .644$), and may be labeled as an Externalizing grouping. The second factor was defined by high loadings for Anxious ($M = .883$) and moderate loadings for Withdrawn/Depressed ($M = .461$). This factor may be labeled as an Internalizing grouping. The largest cross-loading was for Withdrawn/Depressed on the Externalizing grouping ($M = .272$). The mean loading of Sleep Problems or Somatic Problems never exceeded .294 on either the Externalizing or Internalizing dimension.

The Pearson correlations between Internalizing and Externalizing raw scores were .39, .40, and .53 in the clinical, community, and twin sample, respectively. This reflects the fact that children who have elevated scores in one of the two areas also tend to have somewhat elevated scores in the other area. Test-retest rs for the Internalizing and Externalizing groupings were .81 and .90, respectively. Parent-parent correlations for the Internalizing grouping were .48 in the combined clinical and community sample, and .69 in the twin sample. Parent-parent correlations for the Externalizing grouping were .66 in the combined clinical and community sample, and .75 in the twin sample.

Discussion

The study reported in this paper showed that the CBCL/2-3 enables a distinction between different dimensions of problem behavior in children 2-3 years of age as reported by parents. Seven syndromes were identified: Oppositional, Withdrawn / Depressed, Aggressive, Anxious, Overactive, Sleep Problems, and Somatic Problems. Congruity found for the factor solutions across three independent samples made it possible to derive similar syndromes and scales.

Although approximately the same factor structure was found in three independent Dutch samples, there were differences compared to the American factor structure (Achenbach, 1992). The Dutch scales, constituting an Externalizing grouping, were labeled Oppositional, Aggressive, and Overactive. The American Externalizing grouping consisted of an Aggressive Behavior scale and a Destructive Behavior scale.

The Oppositional scale is most concordant with the American Aggressive Behavior scale composed by Achenbach (1992), but three out of four items with the highest loadings on the American Aggressive Behavior scale are not in the Oppositional scale. Therefore another label was used for this scale. A label that is broad enough and yet distinguishing the behavior from hostile aggression and conduct disorder is "o-ppositional". The DSM-III-R (APA, 1987) and ICD-10 (WHO, 1989) diagnostic guidelines for Oppositional Defiant Disorder include many of the behaviors covered by the Oppositional syndrome.

The Dutch Aggressive scale contains equal numbers of items with high loadings on the American Aggressive and Destructive scales. The Dutch Aggressive syndrome is represented by items that almost all reflect aggressive acts to individuals (including animals) and properties. These behaviors represent the core of many definitions of aggression (see Parke & Slaby, 1983). Therefore, the label "Aggression" was considered the most suitable label for the scale derived in the present study.

The Dutch Overactive scale had no American counterpart in Achenbach's (1992) study. The Overactive syndrome items reflect concentration problems, short attention span, and overactivity. A hyperactivity-inattentiveness factor has only once been found for preschoolers (Behar & Stringfield, 1974), but repeatedly for children from age 6 onwards using parent as well as teacher ratings (see Taylor, 1988). In accordance with this research, the Dutch scale was labeled Overactive.

The American Internalizing grouping includes an Anxious/Depressed syndrome and a Withdrawn syndrome. A comparison between the 1987 and 1992 scales indicated that the invariance of the components Withdrawn and Anxious/Depressed was poor across samples. The Dutch scales constituting the Internalizing grouping were labeled Anxious and Withdrawn/Depressed. Compared to the American scales, this suggested a somewhat different grouping of the same items. Like for the American scales, the Dutch Withdrawn/Depressed scale appeared to be the least stable scale. These findings suggested that it is most difficult to obtain a stable differentiation between young preschoolers' Internalizing behaviors.

Sleep problems, defined by various behaviors including sleep disturbance and resistance to go to sleep constitute a well-known problem to many of the parents of young children as well as to clinicians and consultants. The Sleep Problems scale was perfectly replicated across cultures. The American version of the Somatic Problems scale included 14 items. The Dutch version of the Somatic Problems scale was comprised of only three items. Five items that loaded $\geq .30$ in at least one of the samples, but given the inclusion criteria only 3 were retained. One reason for the relatively small Dutch Somatic Problems

scale may be that several of the items on somatic problems in the checklist ask for complaints without medical cause. Dutch respondents may have been hesitant to score these items because they felt uncertain about the medical nature of the complaints. This is indicated by the low frequency of somatic problems items in the Dutch sample compared to the American sample.

Both the factor intercorrelations and the two-factor model showed a clustering of the Oppositional, Aggressive, and Overactive syndromes, and a clustering of the Anxious and Withdrawn/Depressed syndromes. These clusters are perhaps best interpreted as Externalizing and Internalizing groupings (cf. Achenbach, 1992). The χ^2 tests suggested that the relations among the syndromes could not entirely be explained by two higher-order factors. The terminology of groupings also accounts for relations between certain syndromes, but does not assume a specific factor model.

Compared to analyses on American samples, analyses of CBCL/2-3 data in three independent Dutch samples suggested a somewhat different factor structure. However, the present analyses on Dutch samples supported American findings, which showed that the CBCL/2-3 allows a further differentiation between problem behavior in preschool children than the broad-band groupings found for teacher ratings. A study on the discriminative and predictive validity of the scales obtained in these analyses is now in progress.

1. Data obtained for pairs of twins are not independent. As a consequence, standard errors and tests of significance are incorrect, and should not be used in analyses when data from both twins are included. Estimates of the factor loadings, however, are correct (see Goldstein, 1987, p. 30; Weng, 1990, pp. 28-31). We fitted the factor model separately to maternal ratings of the first twin, maternal ratings of the second twin, paternal ratings of the first twin, paternal ratings of the second twin, maternal ratings of both twins, and paternal ratings of both twins. Of the 101 loadings estimated from each of the analyses of maternal and paternal ratings in the sample including both twins, 97 and 98 loadings, respectively, were within the range of loadings obtained from each of the analyses of maternal and paternal ratings in the samples including only one member of the twin pair.

By computing the mean of the parental ratings, it was assumed that mothers and fathers observe the child's behavior in similar situations, and that they share a common understanding of the behavioral descriptions. Loadings obtained from the separate analyses of the maternal and paternal ratings of both twins were highly correlated with the loadings obtained from the analyses using the mean of maternal and paternal ratings (r = .996 and .992, respectively). If only paternal ratings had been used, one additional item (15. Disobedient) would have been included in the Oppositional scale, because the cross-correlation of this item on the Aggressive scale was below .30 using paternal ratings. Use of only maternal ratings would have had no effect on the composition of the scales. Other analyses of the twin data also yielded negligable differences in means and variances of the maternal and paternal ratings (Van Den Oord, Koot, Boomsma, Verhulst, & Orlebeke, submitted), and provided support for the assumption stated above (Van Den Oord, Verhulst, Boomsma, & Orlebeke, submitted).

6
A Twin-Singleton Comparison of Problem Behavior in 2-3-Year Olds

Edwin J.C.G. Van Den Oord[1,2], Hans M. Koot, Dorret I. Boomsma[2], Frank C. Verhulst[1], J.F. Orlebeke[2].

Abstract

Twin-singleton differences in problem behaviors in 2-3-year olds were studied. Maternal ratings of childrens' problem behaviors were obtained with the CBCL/2-3. The twin sample consisted of 1281 twin pairs (407 MZ, 874 DZ), the singleton sample consisted of 420 children from the general population. Results indicated that the general level of problem behaviors in twins was broadly comparable to that in singletons. Five of the seven scales showed lower scores for DZ twins versus MZ twins and singletons. However, these differences were small. Standard deviations for 2 of the 7 scales were somewhat smaller for singletons than for twins. Higher means for boys were found for the total problem score, and the Aggressive and Overactive syndromes.

Introduction

Twin studies are frequently used to study genetic influences on problem behaviors in children. However, the generalization of findings from twin studies to the general population may be limited by differences between twin and non-twin samples (Gau, Silberg, Erickson, & Hewitt, 1992).

Rutter, and Redshaw (1991) discussed several reasons for possible differences between twins and singletons. Firstly, there are various biological differences. For example, twins have a higher rate of congenital anomalies, and also suffer from a higher rate of obstetric and perinatal complications. Other examples are the lower birth weight, and the shorter lenght of gestation in twins (Bulmer, 1970, p. 46).

A second set of reasons for twin-singleton differences is associated with the upbringing and life experiences of twins (Rutter, & Redshaw, 1991). Parent-child interactions may be different for twins versus non-twins, because parents of twins have to divide their resources between two children of a comparable developmental level. Interactions between the twins themselves might also be a source of twin-singleton differences.

There are a limited number of studies that have compared problem behaviors in twins and singletons in middle childhood and adolescence. Most studies showed only small differences, and suggested that the level of problem behavior is broadly comparable (Ghodsian, 1989, see Rutter, & Redshaw, 1991; Golding, & Osborn, 1989 see Rutter, & Redshaw, 1991; Hay, & O'Brien, 1984, 1987). In contrast, a recent study by Gau, Silberg, Erickson, and Hewitt (1992) showed small but consistently higher levels of

[1] Department of Child and Adolescent Psychiatry, Sophia Children's Hospital-Erasmus University Rotterdam, the Netherlands

[2] Department of Psychonomics, Free University, Amsterdam, the Netherlands

problem behaviors in twins.

For children of preschool age, even less is known about the level of problem behavior in twin versus non-twin samples. Kim, Dales, Connor, Walters, and Witherspoon (1969), studying 13 pairs of monozygotic twins and 22 singletons, found lower levels of aggressive behaviors in twins than in singletons. Lytton, Conway, and Sauvé (1977), and Lytton (1980, p. 157) reported for a sample of 46 twin pairs and 44 singletons, lower rates of compliance with parental requests in twins. The small number of subjects and the limited range of problem behaviors that were adressed in these two studies, make it difficult to draw firm conclusions concerning the level of problem behaviors in twins and singletons.

The aim of the present paper was to study twin-singleton differences in problem behaviors in 2-3-year old children. Maternal ratings of problem behaviors in twins (1281 pairs) and singletons (420 children) were obtained with the Child Behavior Checklist for Ages 2 to 3 (Achenbach, 1992). Mean problem scores and standard deviations were compared for groups of monozygotic twins (MZ), dizygotic twins (DZ), and singletons.

Methods

Measure

The Child Behavior Checklist for Ages 2 to 3 (CBCL/2-3; Achenbach, 1992), is an assessment instrument to obtain parental ratings of problem behaviors in 2-3-year olds. The CBCL/2-3 was modeled on the CBCL for ages 4-18 (Achenbach, 1991a). It consists of 99 items describing a broad range of problems. Parents are requested to circle a 0 if the problem is not true of a child, a 1 if the item is somewhat or sometimes true, and a 2 if it is very true or often true.

Dutch syndromes for the CBCL/2-3 were derived by Koot, Van Den Oord, Verhulst, and Boomsma (submitted). The Dutch syndromes differ somewhat from those reported for American samples (Achenbach, 1991b), and are labeled Oppositional, Withdrawn/Depressed, Aggressive Behavior, Anxious, Overactive, Sleep Problems, and Somatic Problems. Somatic Problems was not studied in the present paper because it could not be reliably assessed, and frequencies of problems comprising this syndrome were low in the twin sample and community sample. The total problem score was also studied in this paper. The total problem score is the sum of all 99 items, and includes items that do not appear in one of the syndrome scales.

Subjects

Twin Sample. In the Netherlands, about 85% of the parents of all newborns are paid a home visit by a commercial organization. During this home visit parents of twins are asked to participate in the Dutch Twin Register kept by the Department of Psychonomics of the Free University of Amsterdam. Between 40% and 50% of all multiple births in the Netherlands are registered.

Questionnaires were mailed to 1792 parents of 3-year-old twins. Non-responders were sent reminders and, when no response was obtained, contacted by phone. Completed questionnaires were returned by 1306 parents (73%).

For 223 same-sex twin pairs results from a blood test were available to determine the zygosity of the twins. This test was based on an analysis of 26 blood group polymorphisms. For 1004 twin pairs information about zygosity was obtained from a questionnaire completed by parents when almost all twins were about 2 years old. Forty families indicated that they were not certain about the zygosity of their twin. These parents were

contacted by phone. Twenty-five pairs were discarded because their parents were still uncertain. This procedure left a sample of 218 MZ female, 189 MZ male, 233 DZ female, 252 DZ male, and 389 opposite-sex pairs.

To establish the reliability of the zygosity determination with the questionnaire, blood test results were compared with the zygosity information from the questionnaire. For the 189 same-sex twin pairs for whom both blood test and questionnaire results were available, the agreement was 82,5%. It could very well be that parents who were uncertain about their twins' zygosity were more likely to consent to a blood test. Perhaps, this percentage is therefore better viewed as the lower bound of the reliability of the questionnaire.

Community sample. Subjects in the community sample were 420 children (215 boys, 205 girls) from a target sample of 469 children (90%) randomly selected from the population of the Dutch province of Zuid-Holland (for a full description of the community sample, see Koot, & Verhulst, 1991). First, a letter was sent to the parents of the 469 eligible children explaining the purpose of the study, the way in which the child was selected, and an announcement that an interviewer would contact them. Then the parents were contacted by telephone and visited by one of four trained female interviewers, who had an education at the master's level in special education or psychology. The interviewer read the CBCL/2-3 problem items regarding the target child aloud, and scored the parent's responses.

Demographic characteristics of both samples are shown in Table 1.

Table 1. Demographic characteristics of the twin sample and the community sample.

	twin Sample		community sample
	monozygotic	dizygotic	
	p=407	p=874	n=420
Age child (in months)			
mother	41.9(3.96)	42.1(4.02)	36.4(7.06)
father	41.6(3.43)	41.9(3.78)	
Age of the parents (in years)			
mother	32.5(4.03)	33.2(3.77)	31.5(4.41)
father	35.2(4.94)	35.8(4.47)	34.2(5.04)
Paid labour			
mother	29.2%	28.0%	32%
father	98.3%	97.4%	93%
Level of parental occupation			
mother	3.47(1.32)	3.66(1.39)	3.53(1.41)
father	3.47(1.36)	3.54(1.42)	3.68(1.44)

Note. Standard deviations are in parentheses. The level of parental occupation (ITS, Van Westerlaak, Kropman, & Collaris, 1975) is the mean of a six-step scale (6 = highest level). P is number of pairs, n is number of children.

The twins were 5 to 6 months older than the singletons. Parents of DZ twins were somewhat older than parents of MZ twins, and parents of MZ twins were somewhat older than parents of singletons. Compared to the community sample, more fathers but fewer mothers in the twin sample were employed. For those parents who were employed, the level of parental occupation (measured on the 6-step scale of Van Westerlaak, Kropman, & Collaris, 1975) was similar in both samples.

Data analyses

Children in the community sample were mainly rated by their mothers (98.3%). In the twin sample almost all twins were rated by their mothers, and 81% was also rated by their fathers. To study the equivalence of the twin sample and the community sample, the means and standard deviations of the maternal ratings in both samples were compared. LISREL 7 (Jöreskog, & Sörbom, 1989) was used to perform significance tests by a simultaneous analysis of the 7 groups (MZ girls, MZ boys, DZ girls, DZ boys, DZ opposite sex, singleton girls, singleton boys) in this study.

Tests for group differences were performed for means and standard deviations separately. First, we tested whether there were significant differences between twins and singletons. Then, tests for sex differences were employed. If there were significant differences between twins and singletons, we examined whether MZ twins differed from DZ twins, MZ twins differed from singletons, or if DZ twins differed from singletons.

To obtain an impression of the magnitude of a possible difference in means, Cohen's (1988, p. 20) effect size was computed by dividing the absolute difference between the means of the groups by a pooled estimate of the standard deviation. According to Cohen's criteria (1988, p. 40), an effect size of .2 represents a small effect, .5 represents a medium effect, and .8 represents a large effect. To obtain an impression of the magnitude of group differences in standard deviations, the ratio of the standard deviations was computed by dividing the smaller standard deviation by the larger standard deviation.

Data obtained for pairs of twins are not independent. To perform accurate significance tests, it is necessary to account for these dependent observations. In LISREL this can be done by specifying in the groups of twins 2 variables, 1 for each child. LISREL, however, requires that every group has the same number of variables, but in the groups of singletons there is only 1 variable. Therefore, in addition to a variable for the ratings of the singletons (P), a dummy variable D with pseudo values VARIANCE(D)=1, COVARIANCE(P,D)=0, and MEAN(D)=0 was specified in the groups of singletons (analogous to the way missing data can be handled in LISREL, Jöreskog, & Sörbom, 1989 p. 259). For the 7 groups there were 29 (12 variances, 5 covariances, and 12 means) observed statistics and 6 (2 variances, 2 covariances, and 2 means) statistics associated with the dummy variables. The degrees of freedom were adjusted for these dummy statistics (this was done by putting df=-6 on the OU line of the last group).

In a multi-sample analysis (Jöreskog, & Sörbom, 1989, pp. 227-244) the fit of a model that assumes that means or standard deviations are not equal across groups can be compared with the fit of a model that constrains means or standard deviations to be equal across groups. When the fit of the latter model is significantly worse, it may be concluded that there is a significant difference in means or standard deviations. To test whether the decrease in fit of the more restrictive model was significant, the 'chi-square difference test' was used. The difference of the fit indices of both models, the 'chi-squares', has itself a chi-square distribution with the difference between the degrees of freedom (number of restrictions of the model) of both models as it's degrees of freedom.

To test for group differences in means a baseline model was fitted to the data. This baseline model constrained, within the same-sex twin groups, the mean of the first twin to the mean of the second twin. In the opposite-sex DZ group the mean of girls was constrained to be equal to the mean of girls in the same-sex DZ twin group, and the mean of boys in the opposite-sex group was constrained to be equal to the mean of boys in the DZ same-sex group. To test for differences between twins and singletons, the fit of the baseline model was compared with the fit of a model that constrained, for each sex separately, the means of MZ, DZ, and singletons to each other. To test for differences between girls and boys, the fit of the baseline model was compared with the fit of a model that constrained, within the MZ, DZ, and singleton groups, the means of girls and boys to each other.

When there were significant differences between the means of twins and singletons, we examined which of the groups (MZ twins, DZ twins, or singletons) differed from each other. When there were no significant sex differences, sex differences were not allowed in these pairwise comparisons and the fit the model without sex differences was used as the basis to perform the chi-square difference test.

To compute Cohen's effect size, a pooled estimate of the standard deviation was obtained by constraining in LISREL the standard deviations of all groups to each other. In case of sex differences, an average effect size for differences between MZ twins, DZ twins, and singletons, the effect size was computed by constraining means to be equal for both sexes. In case of differences between MZ twins DZ twins, and singletons, an average effect size for differences between girls and boys was computed by constraining means to be equal for twins and singletons.

To test for group differences in standard deviations, the same tests were employed as for the means. However, this time the standard deviations were constrained to be equal across or within groups, and no constraints were imposed on the means.

To approximate normality, logarithmic transformations were performed on the scores for the Depressed/Withdrawn, Aggressive, Anxious, and Sleep Problems syndromes. With maximum likelihood estimation, normal distributions are neccesary to perform accurate significance tests (Muthén, & Kaplan, 1985). For all significance tests, a probability level of .05 was applied.

Results

For 1056 twin pairs, both parents completed one CBCL/2-3 for each child. By multi-sample analyses in the five twin groups, tests for differences between the maternal and paternal ratings in the twin sample were employed. The fit of a model which estimated different means or different standard deviations was compared with the fit of a model that constrained means and standard deviations (within each sex $\times$ zygosity group) to be equal for both raters.

Chi-square difference tests indicated that the mean scores of the maternal ratings were significantly higher for the total problem score, the Oppositional syndrome, and the Aggressive syndrome. The effect sizes were respectively .06, .07, and .10. According to Cohen's criteria, these differences did not reach the level of 'small effects'. Only for the Oppositional syndrome the difference between the standard deviations of the parental ratings was significant. The standard deviation of the paternal ratings was somewhat smaller (the ratio equaled .94).

Table 2 shows the means and standard deviations of the maternal ratings for the 7

Table 2. Means and standard deviations for maternal ratings and results from tests for group differences.

	MZ		same sex DZ		opp. sex DZ	normative sample		
	girls	boys	girls	boys	girls/boys	girls	boys	results from significance tests
	p=218	p=189	p=233	p=252	p=389	n=199	n=214	
Total problem score								
mean	32.3	34.4	27.7	32.7	25.5/28.8	32.3	34.4	girls < boys, d=.26; DZ < MZ=Co, d=.20
S.D.	20.2	19.0	17.2	18.8	16.6/17.6	16.6	17.0	girls=boys ; Co < MZ=DZ, r=.87
Oppositional								
mean	10.6	10.8	9.32	10.1	8.84/9.18	10.8	10.6	girls=boys ; DZ < MZ=Co, d=.20
S.D.	7.19	6.49	6.29	6.82	6.26/6.48	5.97	6.03	girls=boys ; Co < MZ=DZ, r=.91
Depressed/Withdrawn								
mean	1.14	1.19	1.09	1.36	1.01/1.00	1.02	1.26	girls=boys ; MZ=DZ=Co
S.D.	1.39	1.51	1.66	1.82	1.58/1.47	1.54	2.09	girls=boys ; MZ=DZ=Co
Aggressive								
mean	3.06	4.29	2.47	3.96	2.19/3.55	2.61	3.71	girls < boys, d=.45; DZ < MZ , d=.13
S.D.	2.71	3.10	2.40	2.98	2.27/2.95	2.04	2.94	girls < boys, r=.89; Co < MZ , r=.87
Anxious								
mean	4.13	3.51	3.55	3.69	2.98/3.35	3.27	3.26	girls=boys ; MZ > DZ=Co, d=.13
S.D.	3.64	2.91	3.15	3.31	2.92/2.95	2.80	3.03	girls=boys ; MZ=DZ=Co
Overactive								
mean	2.56	3.16	2.47	2.79	2.24/2.69	3.05	3.24	girls < boys, d=.17; DZ < MZ=Co, d=.17
S.D.	2.13	2.31	2.17	2.22	2.18/2.26	2.25	2.60	girls=boys ; MZ=DZ=Co
Sleep Problems								
mean	2.15	1.86	1.69	1.92	1.65/1.41	2.21	2.0	girls=boys ; DZ < MZ=Co, d=.17
S.D.	2.43	2.08	2.17	2.14	2.12/1.92	2.53	2.57	girls=boys ; MZ=DZ=Co

Note. MZ is monozygotic twins, DZ is dizygotic twins, Co is community sample, S.D. is standard deviation, p is number of pairs, n is number of subjects. For instance, DZ < MZ=Co denotes that there were no significant differences between MZ twins and singletons, but that means or standard deviations were significantly lower/smaller for DZ twins. Cohen's effect size is represented by d, r is the ratio of the standard deviations.

groups in the present study, and presents the results from the significance tests. Chi-squares and probabilities of the tests are shown in appendix 5. Means tended to be smaller for DZ twins than for MZ twins and singletons. For some scales, standard deviations tended to be smaller for singletons than for twins. Means were somewhat higher for boys than for girls.

Except for the Withdrawn/Depressed syndrome, differences between the means of twins and singletons were significant. For 5 of the 6 syndromes these differences consisted of lower means for DZ twins. The means for MZ twins and singletons did not differ significantly from each other. For these five scales, the mean effect size was .20. This indicated a small difference between DZ twins versus MZ twins and singletons.

For 4 of the 7 scales there were no differences in standard deviations between MZ twins, DZ twins, and singletons. The total problem score and Oppositional syndrome, showed somwhat larger standard deviations for twins than for singletons. For the Aggressive syndrome, standard deviations were somewhat smaller for singletons than for MZ twins.

Sex differences in mean scores were found for the total problem score, and the Aggressive and Overactive syndromes. Girls obtained lower scores than boys. For the total problem score and Overactive syndrome the effect size was small. For the Aggressive syndrome the effect size was medium. The Aggressive syndrome was the only scale that showed sex differences in standard deviations. The standard deviation for girls was smaller than for boys.

Discussion

Twin-singleton differences in problem behaviors in 2-3-year olds were studied. Results indicated that the level of problem behaviors in twins was broadly comparable to that in singletons. Five of the seven syndromes showed lower scores for DZ twins versus MZ twins and singletons. However, these differences were small. The standard deviations of 2 of the 7 scales were somewhat smaller for singletons than for twins. Higher means for boys were found for the total problem score, and the Aggressive and Overactive syndromes. For the Aggressive syndrome, the standard deviation for girls was somewhat smaller than for boys.

In the presence of equal means for MZ twins and singletons, the somewhat lower scores for DZ twins is hard to explain. Inspection of the data suggested that especially DZ opposite sex twins obtained lower scores. For the total problem score, Aggressive syndrome, and Sleep problems, the difference between same-sex and opposite-sex twins was even significant.

Results from the present study were in agreement with most studies that have compared problem behaviors in twins and singletons in middle childhood and adolescence (Rutter, & Redshaw, 1991). However, they do not agree with the CBCL study by Gau, Silberg, Erickson, and Hewitt (1992), which showed small but consistently higher levels of problem behaviors in twins than in singletons. Gau, Silberg, Erickson, and Hewitt (1992) compared maternal ratings of problem behaviors in 1824 twins with a American normative sample, which consisted of a community sample with the exclusion of children who recently had received mental health services (Achenbach, 1983). A number of reasons such as the age difference, the use of an American versus a Dutch sample, different respons rates (44% versus 73% in the present study), and the use of a normative sample versus a community sample in the present study may have contributed to the

different findings.

The present study showed few differences in the distributions of problem scores in 2-3-year old twins and singletons. These findings provided support for the generalizability of findings from twin studies of problem behaviors in preschool children, to the general population.

7

A Genetic Study of Maternal and Paternal Ratings of Problem Behaviors in Three-Year-Old Twins

Edwin J.C.G. Van Den Oord[1,2], Frank C. Verhulst[1], Dorret I. Boomsma[2], J.F. Orlebeke[2].

Abstract

Genetic and environmental influences on problem behaviors in three-year-old twins were studied. Fathers' and mothers' ratings of problem behaviors in twins (218 MZ female, 189 MZ male, 233 DZ female, 252 DZ male, and 389 DZ opposite sex pairs) were obtained with the Child Behavior Checklist for Ages 2-3. Model fit indices indicated that mothers and fathers assessed similar behaviors in their children. Genetic influences accounted on average for 65% of the trait variance. Shared environmental influences accounted on average for 12%, and non-shared environmental influences for 21% of the trait variance. Sex differences in genetic and environmental influences on problem behaviors were small. Evidence for sibling contrast effects was found for the Anxious and Overactive syndromes.

Introduction

There is a paucity of research on genetic influences on psychopathology in preschool children. A number of studies have demonstrated the importance of genetic factors in problem behaviors in older children and adolescents (Rutter et al., 1990). However, little is known about genetic influences on problem behaviors in preschool children.

With the possible exception of infantile autism, most problem behaviors in preschool children do not form clearcut diagnostic categories. Problem behaviors in children generally involve quantitative variations of behavior that most children display to some degree. It is therefore preferable to examine genetic influences on psychopathology assessed as quantitative variations of behavior rather than all-or-none categories. From a genetic point of view it is likely that for these continuous variations the effects of many genes are involved (McGuffin, & Gottesman, 1985), and that methods from quantitative genetic theory have to be applied for studying child psychopathology.

For the assessment of problem behaviors in preschool children, parents are a primary source of information. A meta-analysis by Achenbach, McConaughy, and Howell (1987) showed that the mean correlation between mothers' and fathers' ratings of problem behaviors in the same child is about .6. Several authors have attributed the moderate parental agreement to properties of the measurement instrument or rater (Bates, Freeland,

[1] Department of Child and Adolescent Psychiatry, Sophia Children's Hospital-Erasmus University Rotterdam, the Netherlands

[2] Department of Psychonomics, Free University, Amsterdam, the Netherlands

& Lounsbury, 1979; Neale, & Stevenson, 1989). Both unreliability and a tendency of an individual rater to overestimate or underestimate scores consistently, attenuate the agreement between different raters. However, mothers and fathers may interact differently with the same child, and see the child in somewhat different situations. Imperfect mother-father agreement may therefore indicate that target variables differ for each of the two raters, rather than that their reports are biased or unreliable.

Genetic and environmental influences on problem behaviors in preschool twins were studied in the present paper. Fathers' and mothers' ratings of problem behaviors in their twins (218 MZ female, 189 MZ male, 233 DZ female, 252 DZ male, and 389 DZ opposite sex pairs) were obtained with the Child Behavior Checklist for Ages 2-3 (CBCL/2-3, Achenbach, 1992). This paper is one of the first reports on genetic influences on problem behaviors in preschool children. To obtain accurate heritability estimates it is neccesary to apply corrections for rater bias and unreliability. These corrections are not possible with ratings of a single rater. Ratings of both parents were therefore obtained. Corrections for rater bias and unreliability can be applied under the assumption that parents assess the same behaviors. We therefore first evaluated whether mothers and fathers report on the same or different behaviors.

Method

Sample

In the Netherlands, about 85% of the parents of all newborns are paid a home visit by a commercial organization. During this home visit parents of twins are asked to participate in the Dutch Twin Register (NTR) kept by the department of Psychonomics of the Free University in Amsterdam. Forty to 50% of all multiple births in the Netherlands since 1987 are registered.

Questionnaires were mailed to 1792 parents of 3-year-old twins. Non-responders were sent reminders and, when no response was obtained, contacted by phone. Completed questionnaires were returned for 1306 twins (73%).

For 223 same-sex twin pairs results from a blood test were available to determine the zygosity of the twins. This test was based on a comparison of 26 blood group polymorphisms. For 1004 twin pairs information about zygosity was obtained from questionnaire completed by parents when almost all twins were about 2 years old. Forty families indicated that they were not certain about the zygosity of their twin. These parents were contacted by phone. Twenty-five twin pairs were discarded because their parents were still uncertain. This procedure left a sample of 218 MZ female, 189 MZ male, 233 DZ female, 252 DZ male, and 389 opposite sex pairs. The twins' mean age was 42.06 months (standard deviation 4.00).

To establish the reliability of the zygosity determination with the questionnaire, blood test results were compared with the zygosity information from the questionnaire. For the 189 same-sex twin pairs for whom both blood test and questionnaire results were available, the agreement was 82,5%. It could very well be that parents who were uncertain about their twins' zygosity were more likely to consent to a blood test. Perhaps, this percentage is therefore better viewed as the lower bound of the reliability of the questionnaire.

In an earlier paper, the demographic characteristics of the twin sample were presented and twin-singleton differences in problem behaviors was studied (Van Den Oord, Koot, Boomsma, Verhulst, & Orlebeke, submitted). Results indicated that the

general level of problem behaviors in twins was broadly comparable to that of children from a community sample (N=420).

Assessment instrument

The Child Behavior Checklist for children aged 2 to 3 (CBCL/2-3, Achenbach, 1992) is a rating scale for assessing behavioral/emotional problems in 2-3-year-old children. The CBCL/2-3 was modeled on the CBCL for ages 4-18 (Achenbach, 1991). It consists of 99 items describing a broad range of problems. Parents are requested to circle a 0 if the problem is not true of a child, a 1 if the item is somewhat or sometimes true, and a 2 if it is very true or often true.

Dutch syndromes for the CBCL/2-3 were derived by Koot, & Van Den Oord, Verhulst, and Boomsma (submitted) through applying item analyses on a clinical sample, a community sample, and the twin sample from the present study. The analyses yielded seven syndromes which were labeled Oppositional, Withdrawn/Depressed, Aggressive Behavior, Anxious, Overactive, Sleep Problems, and Somatic Problems. The syndromes differed somewhat from those reported for American samples (Achenbach, 1992). Somatic Problems was not studied in the present paper because it could not be reliably assessed, and frequencies of problems comprising this syndrome were low. The total problem score, which is the sum of the scores for the 99 problem items, was also studied. It includes items that do not appear in the syndrome scales.

Model

Hewitt, Silberg, Neale, Eaves, and Erickson (1992) discussed the application of three classes of models for the joint analysis of mothers' and fathers' ratings of twins. The first model is the biometric model (depicted in Figure 1 for same-sex pairs). The biometric model assumes that mothers and fathers assess different, but possibly correlated, behaviors. This model may be appropriate if mothers and fathers observe the child's behavior in distinct situations, or if they do not share a common understanding of the behavioral descriptions. Within this model disagreement between raters is not automatically regarded as error. Instead, it recognizes that each informant may, from his own perspective, provide different but valid information on the children's functioning. Both the maternal ratings (MRT) and paternal ratings (FRT) are decomposed in an additive genetic (A), a shared environmental (C), and a non-shared environmental (E) factor. Subscripts m and f are used to distinquish the factors which underly the maternal (m) or paternal (f) ratings. Parameters h (additive genetic effect), c (shared environmental effect), and e (non-shared environmental effect) can be viewed as factor loadings or regression coefficients. Parameters subscripted m refer to loadings from maternal ratings of twin 1 (MRT1) or twin 2 (MRT2), on the factors of the maternal ratings. Parameters with subscript f refer to loadings from paternal ratings (FRT). Behaviors assessed by mothers and fathers are allowed to correlate. This correlation is decomposed in an additive genetic, a shared environmental, and a non-shared environmental contribution. Parameters subscripted fm refer to genetic and environmental contributions to the correlation between behaviors assessed by each rater.

In contrast to the biometric model, the bias model (depicted in Figure 2 for same-sex pairs) assumes that both parents assess exactly the same behavior and share a common understanding of the behavioral descriptions. Disagreement between raters is regarded as error. This error occurs because of rater bias, and unreliability. Rater bias is the tendency of an individual rater to overestimate or underestimate scores consistently compared to the mean of the raters. Examples of causes of rater bias are stereotyping, response styles in filling out questionnaires (like avoiding extreme categories), or different

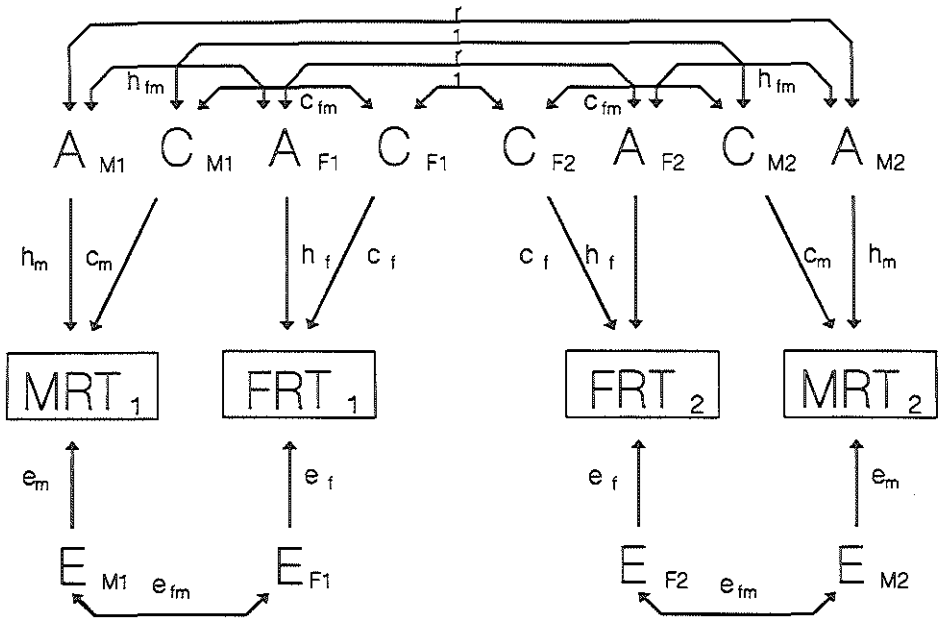


Figure 1. Biometric model.

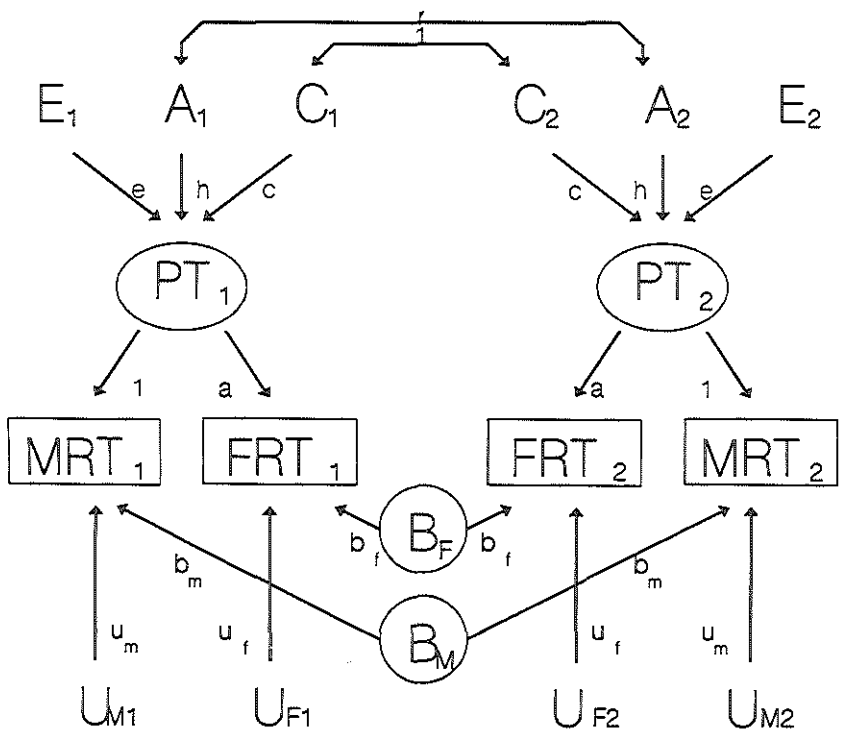


Figure 2. Bias model.

normative standards employed by different raters (Judd, Smith, & Kidder, 1991). Factors, A, C, E, and parameters h, c, e do not depend on the rater and are not subscripted. The behavior assessed by both raters is denoted PT_1 for twin 1, and PT_2 for twin 2. The variance of PT_1 and PT_2 is called trait variance because it concerns the behavior of the child. Parameter b represents the effect from bias factor B on the parental ratings. It is subscripted m and f because the magnitude of the rater bias is allowed to depend on the rater. The unreliability of the measurement is represented by factor U. Its effect u was allowed to depend on the type of rater. The variance due to rater bias and unreliability may be called unique variance because it is unique to each type of rater. The bias model acknowledges that the magnitude of the trait variance does not need to be equal for both raters. This is achieved by estimating for one rater the loading from the observed ratings on the child's behavior (PT). This loading, denoted a, has the same interpretation as a factor loading in factor analysis. In figure 2 we fixed the loading from the mother ratings at one, the loading from the paternal ratings was left free.

The psychometric model (depicted in Figure 3 for same-sex pairs) is a combination of the biometric model and the bias model. It assumes that both parents partially assess the same behavior. In addition there is a component which is unique to each rater.

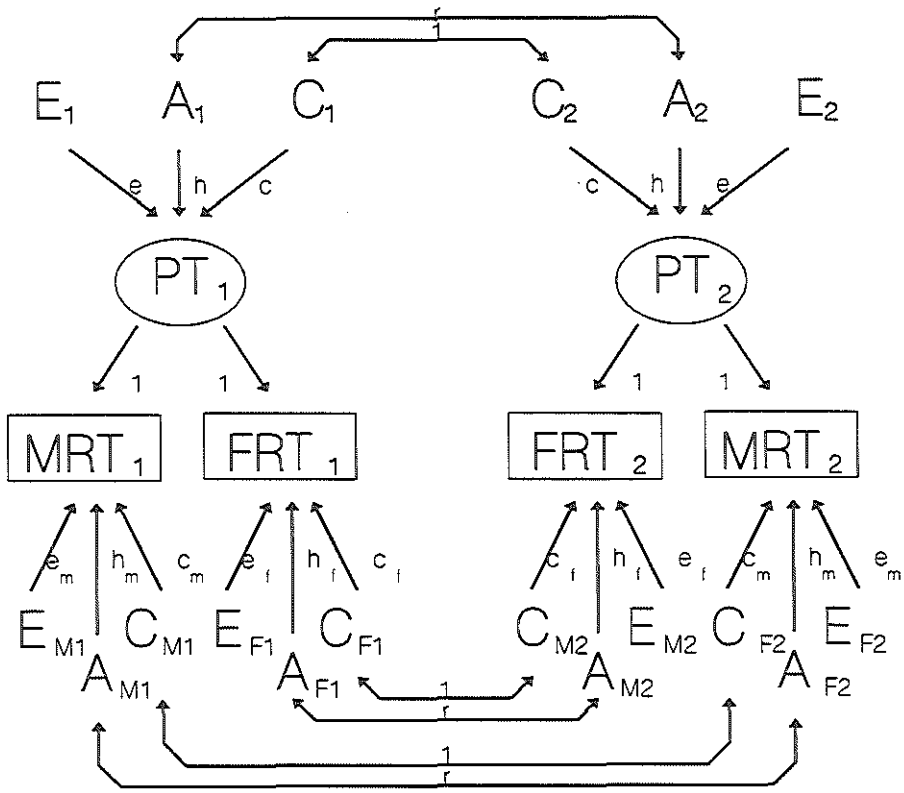


Figure 3. Psychometric model.

In the psychometric model, genetic and environmental factors influence the common part (denoted PT), and the unique part of the parental ratings. Factors and loadings not subscripted m or f concern the common part. Similar to the bias model, the variance of the common part may be called trait variance. Factors and loadings subscripted m and f are associated with the unique part, and contribute to the unique variance.

In this study, all three models allowed 'specific scalar sex limitation' (Heath, Neale, Hewitt, Eaves, & Fulker, 1989; Neale, & Martin, 1989). A model with specific scalar sex limitation assumes that the same genes and environments determine behavior in girls and boys, but that their effects depend on the sex of the child. A consequence of a model with specific scalar sex limitation is that the relative importance of each factor, for instance the heritability, may be different for girls and boys.

Model fitting

The computer program LISREL 7 (Jöreskog, & Sörbom, 1989) was used to analyze the data through a simultaneous analysis of the variances/covariances in the five Sex × Zygosity groups. The implementation of the biometric and psychometric model in LISREL can be achieved by approaches illustrated by Heath, Neale, Hewitt, Eaves, and Fulker (1989) and McArdle and Goldsmith (1990). The implementation of the bias model has been illustrated by Hewitt, Silberg, Neale, Eaves, and Erickson (1992).

Logarithmic transformations were performed on the scores for the Depressed/Withdrawn, Aggressive, Anxious, and Sleep Problems syndromes to approximate normality. With maximum likelihood estimation, normal distributions are neccesary to perform accurate significance tests (Muthén, & Kaplan, 1985). For all significance tests, a probability level of .05 was applied.

Results

For 1056 twins, both parents completed one CBCL/2-3 for each child. For 10 twin pairs only fathers' ratings were available, and for 235 twin pairs only mothers' ratings were available. For 4 twins, one child was rated by the mother and the other by the father, and for one twin pair only one child was rated. For 16 twin pairs there were at least two months between completion by the mother and the father. For these twins only the ratings of the fathers were used. Data from 31 families were discarded, because the father and the mother completed the questionnaire together.

Imput matrices were computed under pairwise deletion. For each Sex × Zygosity group, the average of the pairwise sample sizes was used as the number of observations. Input matrices with observed variances, covariances, and correlations for log-transformed and untransformed scales are shown in appendix 6. A larger resemblance between MZ twin pairs compared to DZ twin pairs is indicative of genetic influences. For the Oppositional, Withdrawn/Depressed, Aggressive Behavior, Anxious, Overactive, Sleep Problems, and Somatic Problems, the twin correlations (averaged for mothers and fathers) was for MZ twins respectively .83, .75, .67, .78, .65, .47, .68. For DZ twins the twin correlations (averaged for mothers and fathers) were respectively .66, .42, .40, .47, .27, .08, .41. All scales showed larger MZ than DZ twin correlations. There were near zero, and in some cases negative, DZ twin correlations for the Overactive syndrome.

In the biometric, bias, and psychometric model several of the observed variances and covariances are replicate estimates of the same expectation. By constraining these variances and covariances to each other, within and across zygosity groups, tests can be employed in LISREL which alert us to the way data may depart from the expectations of

the model.

All three models predict that within the four same-sex groups, the variance of maternal and paternal ratings of twin 1 equals the variance of maternal and paternal ratings of twin 2. The covariance between maternal and paternal ratings of twin 1 should equal the covariance between maternal and paternal ratings of twin 2. Finally, the covariance between maternal ratings of twin 1 and paternal ratings of twin 2 has to equal the covariance between maternal ratings of twin 2 and paternal ratings of twin 1. The model which imposes these restrictions on the imput matrices is denoted model 1 in Table 1. Model 2 in Table 1 constrains, in addition, the variances of the maternal and paternal ratings of MZ and DZ twins to each other. Model 2 is a test for the homogeneity of variances across zygosity groups. This model may fail for a variety of reasons (Falconer, 1989). For instance, there may be group differences between MZ and DZ twins because the frequency of DZ twinning is influenced by genetic factors including racial differences, whereas the frequency of MZ twinning is little, if at all, influenced by genetic or racial factors (Falconer, 1989). Sibling interaction, for instance siblings imitating each others behaviors, also gives rise to heterogeneity of variances across zygosity groups (Carey, 1986; Eaves, 1976; Neale, & Cardon, 1992). Model 3 constrains, across zygosity groups,

Tabel 1. Chi-squares obtained from various baseline models.

	model 1.	model 2.	model 3.	biometric	psychom.	bias
	df=16	df=24	df=28	df=32	df=32	df=34
Total problem score						
	20.43	23.26	44.52	56.43	55.62	113.81
	(.201)	(.505)	(.025)	(.005)	(.006)	(.000)
Oppositional						
	24.58	41.17	64.95	71.25	75.29	74.72
	(.078)	(.016)	(.000)	(.000)	(.000)	(.000)
Depressed/Withdrawn						
	27.91	40.74	42.59	49.50	50.62	69.36
	(.032)	(.018)	(.038)	(.025)	(.019)	(.000)
Aggressive						
	18.07	22.58	27.15	28.86	28.37	49.83
	(.320)	(.544)	(.464)	(.626)	(.651)	(.039)
Anxious						
	10.40	23.03	33.48	42.31	44.34	49.75
	(.845)	(.518)	(.219)	(.105)	(.072)	(.040)
Overactive						
	20.27	23.01	35.01	64.70	64.41	71.36
	(.208)	(.519)	(.170)	(.001)	(.001)	(.000)
Sleep Problems						
	12.60	24.19	33.11	37.19	38.66	56.65
	(.702)	(.451)	(.232)	(.242)	(.194)	(.009)

Note. Model 1 equals, within groups, replicate estimates of a particular expectation to each other. In addition Model 2 equals replicate estimates of variances across groups. Model 3 also equals replicate estimates of covariances across groups. Biometric, psychometric, and bias model are explained in text. Chi-squares that are underlined denote a selected model. Probabilities are in parenthesis, d.f. denotes degrees of freedom.

also the covariance between maternal and paternal ratings of the same child to each other. A reason for model failure is, for example, different parental agreement for MZ compared to DZ twins. The fit of the models constraining replicate estimates of the same expectation to each other, the biometric model, the psychometric model, and the bias model, are reported in Table 1 (all models allowed sex differences). Parameter estimates for the biometric, psychometric, and bias model are shown in appendix 7.

Because model 3 is a submodel of model 2 and both these models are submodels of model 1, the difference of the 'chi-squares' (which itself has a chi-square distribution with the difference in number of estimated parameters as degrees of freedom) can be used to test the significance of the decrease in fit. Some tests indicated that the restrictions in the submodel were not appropriate. For the total problem score the difference between model 2 and 3 was significant. Inspection of the data matrices suggested a higher parental agreement for MZ girls than for DZ girls. For the syndrome Oppositional, the difference between model 2 and model 1, and that between model 3 and model 2 were significant. However, no clear pattern in the way the variances and covariances differed across groups could be detected. For Depressed/Withdrawn, the probability of obtaining that or a more extreme chi-square value was .032 for model 1. This suggested that even the restrictions in model 1 were not appropriate. This model failure was probably caused by chance, because it is hard to explain why observed variances/covariances differed between the first and second born child. For the Anxious and Overactive syndromes, the difference between model 3 and model 2 was significant. No clear pattern could be detected in the way covariances differed across zygosity groups. Further restrictions are imposed on the data by fitting the biometric, the psychometric, and the bias model. These models assume a specific model. When the biometric, the psychometric, and the bias models were fitted, especially the fit for Overactive decreased. For the Overactive syndrome, the models may be too simplistic.

Because the biometric, bias, and psychometric models are not submodels of each other, it is not possible to use the chi-square difference test. The choice which model to prefer was based on fit, parsimony, and the interpretability of the model. The biometric model is less parsimoneous compared to the psychometric model because it does not assume that fathers and mothers assess the same behavior. The psychometric model is less parsimoneous than the bias model because it does not assume that parents assess exactly the same behavior. Table 1 shows that the biometric model did not fit much better than either the psychometric model or the bias model. The psychometric and bias model were preferred because they are more parsimoneous models compared to the biometric model. This implies that both parents assessed the same behavior. For the Oppositional syndrome, the bias model was preferred because it yielded a better fit. For Anxious, the psychometric model fitted somewhat better. However, the bias model was preferred because it is more parsimoneous and more easy to interpret.

Table 1 shows that for some scales, even the preferred model had to be rejected. This indicated that the models may have been too simplistic. It was therefore tested if the fit of the preferred model could be improved by allowing non-scalar sex limitation (Heath, Neale, Hewitt, Eaves, & Fulker, 1989; Neale, & Martin, 1989) or sibling interaction (Carey, 1986; Eaves, 1976; Neale, & Cardon, 1992). Although more complex models were tested, there are also a number of statistical reasons that may have accounted for the rejection of the preferred model. Like other test statistics, chi-squares are affected by sample size. The larger the sample, the greater the probability the model will be rejected (Bentler, & Bonnet, 1980; Marsh, Balla, & McDonald, 1988). Because multiple

tests were performed, the probability of wrongly rejecting at least one model was much larger than .05 (Hays, 1983). Syndrome scores were correlated, and tests were therefore not independent. An atypical sample result, caused by normal sample fluctuations, could therefore have affected the tests for all syndromes.

Models with non-scalar sex limitation assume that different genes or different (shared) environments determine behavior in girls and boys.

In the classical genetic design, sibling resemblance is viewed as caused by the 'passive' sharing of genes and environments. A number of authors have suggested that this passive view may be too simplistic (Carey, 1986; Dunn, 1983; Dunn, & McGuire, 1992; Eaves, 1976; Neale, & Cardon, 1992; Patterson, 1982), because siblings also actively influence each others behaviors. These influences can be incorporated in the model by allowing a direct influence s between each twin's behavior on the behavior of the other twin. In the path diagrams of the psychometric and bias model, this sibling interaction could have been represented by an arrow from PT_1 to PT_2, as well as an arrow from PT_2 to PT_1. When s is positive twins cooperate or imitate each others behaviors. Negative values imply contrast or competition effects. The model assumes that sibling interactions continue until an equilibrium is reached. For an interaction process in equilibrium s will be between -1 and 1. More extreme values of s are not realistic, because it implies infinite scale scores. An example of contrast/competition effects is the case in which dominant behavior of one sibling evokes submissive behavior in the other. This submissive behavior on its turn reinforces the dominant behavior.

Non-scalar sex limitation did not improve the fit of the preferred model for any scale. Problem behaviors in both girls and boys were determined by the same genes and the same shared environments. In contrast, sibling interaction yielded a chi-square of 37.59 for Anxious, and a chi-square of 44.22 for Overactive. Compared to chi-squares of respectively 49.75 and 64.41 for models without sibling interaction, these differences were significant. The negative sign of parameter s implied sibling contrast effects.

To obtain parsimonious models, a variety of more parsimoneous models were fitted to the data. After allowing sibling interaction when appropriate, the preferred models were used as a starting point. These models are denoted model 1 in Table 2. First, the genetic and environmental influences on the child's behavior as assessed by both parents were considered. The fit of a model without sex differences in genetic and environmental influences was compared with the fit of model 1, which allowed scalar sex limitation. When the difference was not significant, the model without sex differences was retained. When the difference was significant, we inspected, for each parameter separately, if it could be constrained to be equal for both sex groups, or fixed at zero. The model that was retained from these tests was denoted model 2 in Table 2. Model 3 imposes additional restrictions on model 2. For the psychometric model, model 3 tests whether the structure of the unique variances is similar for girls and boys. For the bias model, model 3 tests whether rater bias, unreliability, and the factor loading, are equal for both sexes. Model 4 imposes another set of constraints on model 2. In contrast to model 3, it does not constrain parameters to be equal for girls and boys, but constrains parameters which are unique to maternal and paternal ratings to each other. Both model 3 and 4 were tested against model 2. For the psychometric model, model 4 tests whether the structure of the unique variances of maternal and paternal ratings are equal. For the bias model, model 4 tests whether rater bias, unreliability, and the factor loading, are equal for maternal and paternal ratings. Model 5 includes the restrictions of model 3 and 4, and is the most restrictive model. Model 5 was tested against model 2, 3, or 4.

Tabel 2. Chi-squares of models which are respecifications of the preferred model.

	model 1.	model 2.	model 3.	model 4.	model 5.
Total problem score					
	55.62	57.06	127.68	66.47	132.36
d.f.	32	34	40	40	43
	(.006)	(.008)	(.000)	(.005)	(.000)
Oppositional					
	74.72	82.51	89.72	93.93	100.20
d.f.	34	38	43	44	46
	(.000)	(.000)	(.000)	(.000)	(.000)
Depressed/Withdrawn					
	50.62	56.59	67.60	69.04	73.22
d.f.	32	36	42	42	45
	(.019)	(.012)	(.007)	(.007)	(.005)
Aggressive					
	28.37	30.75	45.07	36.74	49.59
d.f.	32	34	40	40	43
	(.651)	(.628)	(.268)	(.618)	(.227)
Anxious					
	37.59	41.09	57.06	44.19	58.61
d.f.	33	36	41	42	44
	(.267)	(.257)	(.060)	(.379)	(.069)
Overactive					
	44.22	46.65	54.45	52.99	59.48
d.f.	31	34	40	40	43
	(.058)	(.073)	(.063)	(.082)	(.048)
Sleep Problems					
	38.66	42.71	43.50	45.86	46.56
d.f.	32	35	41	41	44
	(.194)	(.174)	(.365)	(.278)	(.368)

Note. Chi-squares that are underlined denote a preferred model. Probabilities are in parenthesis, d.f. denotes degrees of freedom. Model 1 is the selected model, model 2 constrains parameters for which non-significant sex differences were found to each other, model 3 assumes that the unique variances are equal for girls and boys, model 4 assumes an equal structure of the unique part of the maternal and paternal ratings, model 5 includes the restrictions of model 3 and model 4.

For the Oppositional, Depressed/Withdrawn, and Anxious syndromes there wereno sex differences in genetic and environmental influences on behavior as assessed by both parents. Sex differences for the other scales were small and did not involve more than one parameter. For 4 of the 7 scales there were no sex differences in the structure of the unique variances (when the psychometric model was appropriate), or the properties of the ratings (when the bias model was appropriate). The properties of maternal and paternal ratings were similar for all syndromes, except for the Oppositional syndrome.

Table 3 shows parameter estimates obtained from fitting the preferred model, and the model that resulted from simplifying this model. For Anxious and Overactive

Tabel 3. Parameter estimates from fitting the most parsimoneous best fitting model.

Psychometric model

	h	c	e	s	h_m	h_f	c_m	c_f	e_m	e_f
Total score										
girls	.287	.257	.122	-	.082	.032	.234	.183	.136	.153
boys	.223	.315	.178	-	.146	.199	.174	.103	.183	.059
girls	*.266*	*.280*	.126	-	.053	.053	.211	.211	.146	.146
boys	*.266*	*.280*	.173	-	.176	.176	.143	.143	.072	.072
Depressed/Withdrawn										
girls	.354	.066	.199	-	.295	-.102	.203	.309	.293	.283
boys	.414	.059	.208	-	.193	.253	.222	.188	.323	.255
girls	*.395*	-	*.194*	-	*.251*	*.251*	*.201*	*.201*	*.292*	*.292*
boys	*.395*	-	*.194*	-	*.251*	*.251*	*.201*	*.201*	*.292*	*.292*
Aggressive										
girls	.154	.110	.083	-	.105	.102	.074	.031	.089	.093
boys	.197	.089	.100	-	.066	.125	.113	-.021	.125	.099
girls	.156	*.101*	*.090*	-	.111	.111	.044	.044	.089	.089
boys	.197	*.101*	*.090*	-	.097	.097	.080	.080	.106	.106
Overactive										
girls	1.059	-.347	1.344	-	.793	.605	.515	.059	.837	.987
boys	1.054	.567	1.501	-	.605	.496	.395	.177	.958	1.126
girls	1.530	.143	.961	-.174	.650	.610	.642	.134	.874	.976
boys	1.497	.710	1.103	-.174	.477	.448	.510	.286	.971	1.132
girls	*1.503*	-	*1.045*	*-.165*	*.563*	*.563*	*.427*	*.427*	*.992*	*.992*
boys	*1.503*	.712	*1.045*	*-.165*	*.563*	*.563*	*.427*	*.427*	*.992*	*.992*
Sleep Problems										
girls	.233	.276	.219	-	.192	.152	.083	.135	.182	.178
boys	.316	.080	.217	-	.220	.155	.000	.128	.168	.176
girls	*.321*	.178	*.213*	-	*.187*	*.187*	*.093*	*.093*	*.176*	*.176*
boys	*.321*	-	*.213*	-	*.187*	*.187*	*.093*	*.093*	*.176*	*.176*

Bias model

	h	c	e	s	b_m	b_f	u_m	u_f	a
Oppositional									
girls	4.275	1.641	2.372	-	3.204	1.794	2.404	1.942	1.071
boys	5.389	-.120	1.900	-	2.751	1.981	2.348	2.663	.968
girls	*5.002*	-	*2.205*	-	*2.959*	*1.943*	*2.358*	*2.366*	*1.005*
boys	*5.002*	-	*2.205*	-	*2.959*	*1.943*	*2.358*	*2.366*	*1.005*
Anxious									
girls	.193	-.032	.113	-	.083	.119	.103	.109	.904
boys	.195	.038	.105	-	.098	.032	.134	.080	1.222
girls	.209	.032	.092	-.108	.087	.118	.105	.108	.916
boys	.173	.058	.086	-.108	.097	.037	.113	.082	1.211
girls	*.201*	-	*.090*	*-.098*	**.104**	**.104**	**.107**	**.107**	**1.000**
boys	*.201*	-	*.090*	-.098	**.071**	**.071**	**.116**	**.116**	**1.000**

Note. Parameter h is additive genetic effect, c is shared environmental effect, e is non-shared environmental effect, s is sibling effect, b is rater bias effect, and u is the square root of the unreliability. Parameter a is the loading from the paternal rating on the phenotype, the loading from the maternal rating is fixed at one. Subscript m refers to mother, subscript f refers to father. Number in italics denote estimate constrained to be equal for both sexes, number in bolface denote estimate constrained to be equal for both raters, - denote parameter fixed at zero.

Tabel 4. Percentages of trait and unique variance explained by genetic and environmental influences.

Psychometric model

		trait var.	genetic	shared e.	non-shared	unique var.	genetic	shared e.	non-shared
Total score									
Mother and	girls	71%	43%(30)	48%(34)	10%(7)	29%	4%(1)	65%(19)	31%(9)
Father ratings	boys	76%	40%(36)	44%(33)	17%(13)	24%	55%(13)	36%(9)	9%(2)
Depressed/Withdrawn									
Mother and Father ratings	girls/boys	51%	81%(41)	-	19%(10)	49%	33%(17)	21%(11)	45%(22)
Aggressive									
Mother and	girls	66%	57%(38)	24%(16)	19%(13)	34%	55%(19)	9%(3)	36%(12)
Father ratings	boys	68%	68%(46)	18%(12)	14%(10)	32%	35%(11)	24%(8)	42%(13)
Overactive									
Mother and	MZ girls	66%	58%(38)	-	42%(27)	34%	21%(7)	12%(4)	66%(23)
Father ratings	boys	69%	52%(35)	12%(8)	37%(25)	32%	21%(7)	12%(4)	66%(21)
Same-sex	DZ girls	69%	64%(44)	-	37%(25)	31%	21%(7)	12%(4)	66%(20)
	boys	71%	57%(40)	10%(7)	33%(23)	29%	21%(6)	12%(4)	66%(19)
Opposite sex	DZ girls	69%	63%(43)	1%(0)	36%(25)	31%	21%(7)	12%(4)	66%(20)
	boys	72%	55%(39)	14%(10)	31%(23)	28%	21%(6)	12%(4)	66%(19)
Sleep Problems									
Mother and	girls	71%	57%(41)	18%(12)	25%(18)	29%	47%(14)	12%(3)	42%(12)
Father ratings	boys	67	69%(46)	-	31%(20)	34%	47%(16)	12%(4)	42%(14)

Bias model

		trait var.	genetic	shared e.	non-shared	unique var.	bias	unreliability
Oppositional								
Mother ratings	girls/boys	68%	84%(57)	-	16%(11)	32%	61%(20)	39%(13)
Father ratings	girls/boys	76%	84%(64)	-	16%(12)	24%	40%(10)	60%(14)
Anxious								
Mother and	MZ girls	65%	80%(52)	-	20%(13)	35%	49%(17)	51%(17)
Father ratings	MZ boys	70%	80%(55)	-	20%(14)	31%	27%(8)	73%(22)
	DZ girls	68%	82%(55)	-	18%(13)	33%	49%(16)	51%(17)
	DZ boys	71%	82%(58)	-	18%(13)	29%	27%(8)	73%(21)

Note. Percentages in parenthesis are percentages of the total variance. Numbers in boldface are the heritabilities of the trait.

parameter estimates from the models with and without sibling interaction, as well as from the simplified model, are presented.

Genetic influences on problem behaviors as assessed by both parents, were largest. Shared environmental influences were absent for three syndromes. For two syndromes, shared environmental influences were small for one sex, and absent for the other. Evidence for sibling contrast effects was found for Anxious and Overactive. Table 4 presents the percentages of genetic, and environmental variance in the trait, unique variance, and total variance. These percentages were computed on the basis of the simplified model. Illustrations of how these percentages can be computed were by given by Hewitt, Silberg, Neale, Eaves, and Erickson (1992). Sibling interaction influences the trait variance but not the unique variance. Formulas given by Neale, and Cardon (1992) were adjusted to the case of scalar sex limitation to compute the percentages of genetic and environmental variance in the trait variance for models with sibling interaction.

When siblings influence each others behaviors, the parameters which affect the behavior of one child also affect the behavior of the other child. For syndromes with both sex differences and sibling interaction, the percentages differ therefore for opposite and same-sex pairs. For syndromes with sibling effects the trait variance depends on the genetic correlation r. Percentages were therefore different for MZ versus DZ twins.

Except for Depressed/Withdrawn, trait variance constituted about 70% of the total variance. The percentage genetic variance in the trait variance is the heritability of the trait. Genetic influences accounted on the average for 65% of the trait variance. For the Oppositional, Depressed/Withdrawn, and Anxious syndromes, genetic influences accounted for more than 80%. Shared environmental influences accounted on the average for 12% of the trait variance. Non-shared environmental influences accounted for 21%.

Most behavior genetic research involved the ratings of only one rater. The results from these studies can be compared with those from Table 4. With ratings of one rater, observed variances can merely be decomposed in a genetic, a shared environmental, and a non-shared environmental contribution. For the psychometric model, an impression of what the genetic component would have been in an analysis with only one rater, can be obtained by summing the percentages in the total variance of the genetic contribution to the trait variance and the unique variance. The same procedure may be followed for the shared and non-shared environmental components. For the bias model the percentage rater bias in the total variance has to be added to the percentage shared environmental variance in the total variance to obtain an estimate of what the shared environmental component would have been in an analysis with ratings from one rater. The percentage unreliability in the total variance has to be added to the percentage non-shared environmental variance to obtain an estimate of the non-shared environmental component in such an analysis. When the bias model is appropriate, a joint analysis of multiple ratings will always lead to larger heritabilities. In a joint analysis, shared environmental influences are disentangled from rater bias, and non-shared environmental influences from unreliability, thereby increasing the relative importance of genetic influences.

Discussion

Genetic and environmental influences on problem behaviors in three-year-old twins were studied. Models discussed by Hewitt, Silberg, Neale, Eaves, and Erickson (1992) were used to examine whether parents assess the same or different behaviors. A model fitting approach was used to select the best fitting parsimoneous model. This model was

used to estimate genetic and environmental contributions to problem behaviors.

The present paper is one of the first reports on problem behaviors in preschool twins. It is therefore difficult to compare our findings with those from other genetic studies on problem behaviors in preschool children. Several syndromes studied in this paper resemble temperament characteristics or have counterparts in syndromes found for older children. For these syndromes a comparison with other studies can be made.

The Overactive syndrome. The Overactive syndrome resembles temperament characteristics like activity and attention span, and has counterparts in syndromes involving hyperactive and attention problems for older children. Goldsmith & Gottesman (1981) found no significant genetic factor for observer ratings of activity level and attention span at age 4. However, the finding of genetic influence on Overactive is in agreement with results from most studies of activity in preschool twins (Cohen, Dibble, & Grawe, 1977; Neale, & Stevenson, 1989; Plomin, 1986, p. 214). Our results are also in agreement with twin studies (Willerman, 1973; Torgersen, 1982), adoption studies (Safer, 1973; Van Den Oord, Boomsma, & Verhulst, submitted), and family studies (Biederman et al., 1986; Cantwell, 1972; Morrison, & Stewart, 1971; Welner, Welner, Stewart, Palkes, & Wish, 1977), which studied comparable syndromes in older children. For the Overactive syndrome we found a heritability of over .50. This is in agreement with other studies using questionnaire ratings, and does not seem to depend on the age of the subjects (Goodman, & Stevenson, 1989; Plomin, 1986, p. 214; Willerman, 1973).

The Oppositional syndrome. The Oppositional syndrome also provides an opportunity for comparison with findings from other behavior genetic research with preschool children. Components of the Oppositional syndrome, like emotional intensity and negative mood, frequently appear in research on childhood temperament as aspects of the "difficult child" (Lee, & Bates, 1985; Thomas, Chess, & Birch, 1968). The Oppositional syndrome is probably also related to Buss, and Plomin's emotionality scale (Buss, & Plomin, 1975, p.17). Buss, and Plomin (1975) found intraclass correlations of .64(for 81 MZ pairs), and .03(for 57 DZ pairs) for the EASI emotionality scale in a sample of 55 months old twins. Plomin, and Rowe (1979) found, for the same scale, intraclass correlations of .70(36 MZ pairs), and .06(31 DZ pairs) in 43 months old twins,. Neale, and Stevenson (1989) obtained, for the emotionality scale, heritability estimates of 31% for boys and 62% for girls in a sample of 219 MZ and 322 DZ 42 months old twins. Goldsmith, & Gottesman (1981) found a difference in MZ and DZ twin correlation of .28 in observer ratings of irritabilty at the age of four. Matheny, Wilson, Dolan, and Krantz (1981) identified a set of behaviors related to negative aspects of temperament. Several of the behaviors in this set, frequency of temperamental outbursts, crying, demanding attention, and irritability, resemble items of the Oppositional syndrome scale. The mean concordances for these behaviors were 44%(76 MZ pairs), and 25%(44 DZ pairs) at 36 months of age. At 48 months of age the mean concordances were 45%(68 pairs MZ twins)/.27(45 pairs DZ twins). Wilson, Brown, and Matheny (1971) identified a temperament cluster which included five variables: temper frequency; temper intensity; irritability; crying; and demanding attention. When the twins were 3 to 4 years old, the mean concordance for these behaviors was 51% for 95 pairs MZ twins, and 40% for 73 pairs DZ twins 40%. In their study for approximately 68% of the twin pairs (189 MZ, and 315 DZ) results for both co-twins were available. Compared to findings from other studies, results from the present study showed a large heritability (over .80) for the Oppositional syndrome. To some extent this can be attributed to the application of the bias model, with its corrections for rater bias and unreliability.

Other syndromes. For syndromes comparable to syndromes found for older children such as Withdrawn/Depressed, Aggressive Behavior, and Anxious, it is possible to compare our findings with those from genetic studies in older children. Compared to genetic studies in older children, heritabilities found in the present study were somewhat larger (Edelbrock, Rende, & Plomin, 1992; Van Den Oord, Boomsma, & Verhulst, submitted). Partially, this may be attributed to the joint analysis of mothers' and fathers' ratings. An analysis of ratings of one parent would have yielded smaller heritabilities, for the Withdrawn/Depressed and Anxious syndromes.

Mother-father agreement. Both unreliability and rater bias decrease the percentage trait variance of the total variance. For the bias model, and to a lesser extent for the psychometric model, the percentage trait variance may be considered as a measure for the quality of the test. Results from the present study showed that on the average 70% of the variance for the Oppositional and Anxious syndromes, consisted of trait variance. With the exception of Withdrawn/Depressed, the scales for which the psychometric model was appropriate yielded similar percentages. This compares well with a study from Neale, and Stevenson (1989) on temperament in preschool twins, and a study of Hewitt, Silberg, Neale, Eaves, and Erickson (1992) on problem behaviors in 4 to 18 year-old-twins. In addition, a meta-analysis of Achenbach, McConaughy, and Howell (1987) yielded correlations of about .6 between mothers' and fathers' reports of problem behaviors in the same child. Under the psychometric and bias model, which assume that different raters assess the same behaviors, the interparent correlation may be viewed as an estimate of the percentage trait variance. Compared to the .6 reported by Achenbach, McConaughy, and Howell (1987), our results indicated a somewhat larger parental agreement. The relatively high parental agreement may be explained by the age of the subjects. Achenbach, McConaughy, and Howell (1987) found significantly higher interparent correlations for 6-11-year-olds than for adolescents. However, the few studies which included preschool children (Earls, 1989; Field, & Greenberg, 1982) yielded smaller than average interparent correlations. It is possible that parents did not complete the questionnaires independently. In almost 80% of the families, both parents completed the questionnaire on the same day. This may have inflated interparent correlations and spuriously suggested larger parental agreement and better properties of the test. A report from Koot, Van Den Oord, Verhulst, and Boomsma (submitted) provided some support for this hypothesis. For a small sample (N=51) of referred and non-referred children, ratings of both parents assessing the same child were obtained with a procedure which ensured that mothers and fathers rated their child independently. The average mother-father correlation in this sample was .56, which suggests a smaller percentage of trait variance than to the 67% we found.

Sibling contrast effects. Several twin studies of early temperament have shown moderate to high MZ twin correlations accompanied by minimal, and sometimes negative, DZ twin correlations (Neale, & Stevenson, 1989; Plomin, 1986, p. 214; Torgersen, 1982). For instance, the average twin correlation for the EASI Activity scale as reported by Plomin (1986) was .62 for MZ twins, and -.13 for DZ twins. In the present study some of the DZ twin correlations for the Overactive syndrome were negative too. The Overactive syndrome resembles the Activity scale reported by of Plomin (1986). Plomin (1986) explained the negative DZ twin correlations by the mechanism of contrasts effects. Parents might contrast their fraternal twins, and accentuate the existing differences between them. Fraternal twins might also contrast themselves, or as Buss, and Plomin (1984) put it: "One twin partner, who might be slightly more active than the other,

converts this slight edge into a consistent advantage in initiating activities, and the other twin relinquishes the initiative to this patner". The sibling contrast effects found in the present study for the Overactive syndrome may be interpreted in the same way, although in contrast to Plomin (1986) it applied to MZ twins as well. From a technical point of view, it is clear why the model without sibling contrast effects did not fit for the Overactive syndrome. The difference between MZ and DZ twin correlations implied genetic influences. In case of additive genetic influences, the correlation between DZ twins (who also share genetic information) has to be sufficiently large. However, because of the near zero DZ twin correlations the model failed, although LISREL still tried to find an optimal solution by underestimating the genetic effect. When only observed correlations are inspected, genetic influences tended to be overestimated and shared environmental influences underestimated. Sibling contrasts effects reduce both the DZ and MZ twin correlation, with a larger reduction for the DZ twin correlation (see Neale, & Cardon, 1992 for a numerical example).

Genetic explanations, in terms of nonadditive genetic effects, can also account for too low DZ twin correlations. It is therefore difficult establish whether sibling contrast effects or genetic explanations are appropriate. Genetic explanations can explain low DZ twin correlations but cannot account for negative DZ twin correlations. Indeed, for the Overactive syndrome a model with genetic dominance (Heath, Neale, Hewitt, Eaves, & Fulker, 1089) showed a poorer fit than the model allowing sibling effects. Under a model with sibling contrast effects the trait variance is smaller for MZ than for DZ twins (Carey, 1986; Eaves, 1976; Neale, & Cardon, 1992). Genetic explanations do not predict differences in variances across zygosity groups. Covariances between maternal and paternal ratings of the same twin were inspected, because they are under both the psychometric and bias model estimates of the trait variance. Indeed, the chi-square difference test indicated that these covariances were not equal across zygosity groups. For girls the (pooled) trait variances for MZ, DZ same-sex, and DZ opposite sex groups were respectively 2.85, 2.91, and 3.33. This finding suggested that the trait variances deviated in the direction as predicted by sibling contrast effect. For boys the variances were 3.78, 3.54, and 3.77 which is not suggestive of sibling contrast effects.

For the Anxious syndrome we also found evidence of sibling contrast effects. For this syndrome there were too low DZ twin correlations, but no negative DZ twin correlations. The test constraining maternal and paternal ratings of the same twin across groups failed. The pattern of variances for girls .045, .049, and .039, and boys .038, .051, and .042, also provided some support for sibling contrast effects. However, a model allowing genetic dominance was equivalent in fit to the model with sibling interaction. Parameter estimates obtained from fitting this model indicated that dominance effects were even more important than additive genetic influences. From a theoretical point of view large amounts of dominance are not plausible and this model was therefore rejected (Eaves, 1986). However, the large dominance effects could have also arisen by chance because of the high correlation between estimates of additive and estimates of dominance effects when only twin data are used (Eaves, 1986). In addition, other genetic explanations can also account for too low DZ twin correlations.

In conclusion, the results from the present study showed that both parents assessed the same behavior. Genetic influences accounted on the average for 65% of the trait variance. For the Oppositional, Depressed/Withdrawn, and Anxious syndromes, genetic influences accounted for more than 80%. Shared environmental influences accounted on the average for 12% of the trait variance. Non-shared environmental influences accounted

for 21%. Sex differences in genetic and environmental influences on problem behaviors were small. Evidence for sibling contrast effects was found for the Anxious and Overactive syndromes.

8

A Multivariate Genetic Analysis of Problem Behaviors in Three-Year-Old Twins

Edwin J.C.G. Van Den Oord[1], Dorret I. Boomsma[2], Frank C. Verhulst[1], J.F. Orlebeke[2].

Abstract

Multivariate genetic models were fitted to study patterns of problem behaviors in three-year-old twins (446 MZ, and 912 DZ twin pairs). Fathers' and mothers' ratings of problem behaviors were obtained with the Child Behavior Checklist for Ages 2-3. A biometric model with two common genetic, one common shared environmental, and two common non-shared environmental factors fitted almost as well as the saturated unconstrained model for the genetic and environmental covariances. The common non-shared environmental factors produced externalizing/internalizing patterns of problem behaviors. One common genetic factor produced a clustering of the Oppositional, Withdrawn/Depressed, and Overactive syndromes with the Aggressive syndrome. The other common genetic factor produced a clustering of the Oppositional, Withdrawn/Depressed, and Overactive syndromes with the Anxious syndrome. A pattern of similar scores on all dimensions of problem behavior was most suggestive of the common shared environmental factor.

Introduction

Many disordered children show multiple problem behaviors (Caron, & Rutter, 1991; Verhulst, & Van Der Ende, 1993). Overlapping diagnostic criteria, or an incorrect subdivision of syndromes may artificially suggest multiple disorders in the same individual. It is also likely that clinical samples contain a disproportionately large number of children showing multiple problems, because the referral is more likely to be initiated when a child is causing concern in two ways rather than one (Caron, & Rutter, 1991). However, the occurrence of multiple problems should not automatically be viewed as an artefact caused by referral biases or diagnostic flaws. Instead, it may be regarded as an aspect of the complexities of child psychopathology.

On the other hand, the presence of multiple problems does not imply that the child suffers from two or more truly distinct and unrelated disorders. It may be that "comorbid" disorders are different manifestations of the same underlying cause. The co-occurrence of problem behaviors, often in specific patterns, may therefore reflect meaningful higher order syndromes that may be distinguished from other syndromes with respect to prognoses, course, or response to clinical interventions.

---

[1] Department of Child and Adolescent Psychiatry, Sophia Children's Hospital Erasmus University Rotterdam, the Netherlands

[2] Department of Psychonomics, Free University, Amsterdam, the Netherlands

The occurrence of patterns of problem behaviors have frequently been studied with higher order factor/principal components analyses. These procedures have repeatedly identified similar groupings of problem behavior (Achenbach, 1991a p. 63). The groupings generally reflect a distinction between anxious, inhibited behavior on the one hand, and aggressive, antisocial behavior on the other. In child clinical literature the groupings have been designed with terms such as Personality Problem versus Conduct Problem; Internalizing versus Externalizing; Inhibition versus Aggression; and Over-controlled versus Undercontrolled.

Patterns of problem behaviors such as the Internalizing and Externalizing groupings, suggest that different problem behaviors may be variable expressions of the same genetic or environmental cause. By fitting multivariate genetic models to phenotypic variances and covariances, these genetic and environmental causes can be studied. This may, for instance, be useful to distinguish problem behaviors at the level of etiology rather than symptomatology (Plomin, Rende, & Rutter, in press).

The present paper studied genetic and environmental causes of patterns of problem behaviors. Despite the fact that the occurrence of multiple problems is very common in children, the present paper is one of the first multivariate genetic studies of problem behaviors in young children.

Method

Assessment instrument

Fathers' and Mothers' ratings of problem behaviors in their twins were obtained with the Child Behavior Checklist for children aged 2 to 3 (CBCL/2-3; Achenbach, 1991b). The CBCL/2-3 was modeled on the CBCL for ages 4-18 (Achenbach, 1991a). It consists of 99 items describing a broad range of problems. Parents are requested to circle a 0 if the problem is not true of a child, a 1 if the item is somewhat or sometimes true, and a 2 if it is very true or often true.

Dutch syndromes for the CBCL/2-3 were derived by Koot, Van Den Oord, Verhulst, and Boomsma (submitted) through applying item analyses on a clinical sample, a community sample, and the twin sample from the present study. These analyses yielded seven syndromes which were labeled Oppositional, Withdrawn/Depressed, Aggressive Behavior, Anxious, Overactive, Sleep Problems, and Somatic Problems. The Dutch syndromes differed somewhat from those reported for American samples (Achenbach, 1991b). Somatic Problems was not studied in the present paper because it could not be reliably assessed, and frequencies of problems comprising this syndrome were low in the twin sample.

In addition to the syndrome scales an Externalizing and an Internalizing grouping was derived. The Externalizing grouping comprised the Oppositional, Aggressive, and Overactive syndromes. The Internalizing grouping comprised the Anxious, and Withdrawn/Depressed syndromes.

Sample

In the Netherlands, about 85% of the parents of all newborns are paid a home visit by a commercial organization. During this home visit parents of twins are asked to participate in the Dutch Twin Register (NTR) kept by the department of Psychonomics of the Free University of Amsterdam. Between 40 and 50% of all multiple births in the Netherlands are registered.

Questionnaires were mailed to 1792 parents of 3-year-old twins. Non-responders

were sent reminders and, when no response was obtained, contacted by phone. Completed questionnaires were returned by 1377 parents (77%).

For 374 same-sex twin pairs results from a blood test were available to determine the zygosity of the twins. This test was based on a analysis of 26 blood group polymorphisms. For 955 twin pairs information about zygosity was obtained from a questionnaire completed by parents when almost all twins were about 2 years old. Thirty-one families indicated that they were not certain about the zygosity of their twin. These parents were contacted by phone. Nineteen twin pairs were discarded because their parents were still uncertain. This procedure left a sample of 446 MZ, and 912 DZ twin pairs (236 MZ female, 210 MZ male, 238 DZ female, 265 DZ male, and 409 opposite-sex pairs). The twins' mean age was 42.12 months (SD = 3.94).

To establish the reliability of the zygosity determination with the questionnaire, blood test results were compared with the zygosity information on the questionnaire. For the 327 same-sex twin pairs for whom both blood test and questionnaire results were available, the agreement was 81%. It is possible that parents who were uncertain about their twins' zygosity were more likely to consent with a blood test. Perhaps, this percentage is therefore better viewed as a lower bound of the reliability of the questionnaire.

For the greater part, the sample in the present study was used in two other papers to study the representativity of the twin sample (Van Den Oord, Koot, Boomsma, Verhulst, & Orlebeke, submitted) and to study genetic and environmental influences on separate syndrome scales (Van Den Oord, Verhulst, Boomsma & Orlebeke, submitted). Demographic characteristics and results from univariate genetic analyses were reported in these studies. Compared to the previous studies, the sample in the present study comprised 71 additional twin pairs, and for more same-sex twin pairs blood test results were available.

Model

Multivariate genetic models were fitted to study patterns of problem behaviors. Multivariate genetic models make a distinction between genetic and environmental factors that influence all observed variables, and genetic and environmental factors that are unique to each observed variable (Martin, & Eaves, 1977; Boomsma, & Molenaar, 1986). The common genetic and environmental factors explain the covariances between the observed variables. The unique genetic and environmental factors explain the part of the variance that is unique to each variable.

Two classes of multivariate models may be distinguished (Heath, Neale, Hewitt, Eaves, & Fulker, 1989; McArdle, & Goldsmith, 1990; Neale, & Cardon, 1992 pp. 231-259). The two classes differ in the way the common factors influence the observed variables. The way the unique factors influence the observed variables is identical.

The first class of models is usually referred to as "psychometric" or "common pathway" models. Psychometric models assume that genetic and environmental factors influence the observed variables via latent variable(s). The latent variable(s) resemble the higher order factor(s) from an higher order factor analysis. Psychometric models decompose the variance of the higher order factor(s) in genetic and environmental contributions. In psychometric models the pattern in which genetic and environmental factors influence the observed variables is alike, because the common genetic and environmental factors influence the observed variables via higher order factor(s).

The second class of models is usually referred to as "biometric" or "independent pathway" models. In biometric models common genetic and environmental factors directly

influence the observed variables, and there are no intervening higher order factor(s). Consequently, the pattern in which common genetic factor(s) influence the observed variables can be quite different from the pattern environmental factor(s) influence the variables. Compared to the psychometric model, the biometric model is therefore less restrictive.

In the present study, both the common and unique factors consisted of uncorrelated additive genetic (A), shared environmental (C), and non-shared environmental (E) factors. By convention, biometric and psychometric models with one set of common A, C, and E components were labeled "one factor" models. Multivariate genetic models with two sets of common A, C, E components were referred to as "two factor" models.

Model selection

To select the best fitting parsimoneous model, several alternative models were fitted to the data.

First, a baseline model was fitted. The baseline model assumed that there were only unique genetic and environmental factors, and no common genetic and environmental factors. This model specifies zero correlations between the syndromes. The baseline model provided a basis for assessing the improvement in fit which resulted from fitting models which could account for covariances between the syndromes.

The second model also allowed only unique genetic and environmental factors. In contrast to the baseline model, all unique genetic factors were allowed to covary with each other, all shared environmental factors were allowed to covary with each other, and all non-shared environmental factors were allowed to covary with each other. This model was a saturated unconstrained model for the genetic and environmental contributions to covariances between syndromes. This type of model is also referred to as a cholesky or triangular decomposition. This saturated model simply decomposed all observed covariances between the syndromes in a genetic, shared environmental, and non-shared environmental contribution. Psychometric and biometric models are more restrictive compared to this saturated model, because they have to account for the covariances with a few common genetic and environmental factors. Fit indices of the saturated model are therefore upper bounds of the fit indices of psychometric and biometric models.

Finally, one and two factor biometric and psychometric models were fitted to the data. For two factor biometric and psychometric models the problem of factor rotation arises. An orthogonal rotation was specified by leaving the first and second factor uncorrelated and fixing the loading from the Anxious syndrome on the first common factor at zero. This rotation corresponded to the varimax/promax rotation, found for the two factor model for the Dutch syndrome scales (Koot, Van Den Oord, Verhulst, and Boomsma, submitted).

In genetic analyses of separate syndrome scales, no evidence of sex differences was found for 3 of the 6 syndromes (Van Den Oord, Verhulst, Boomsma, & Orlebeke, submitted). Sex differences for the other syndromes were small and never involved more than one parameter. The absence of sex differences suggested that there was no need to complicate the present analyses any further, and sex differences were therefore not studied.

The computer program LISREL 7.20 (extended memory, Jöreskog, & Sörbom, 1989) was used to analyze the observed variances/covariances in the two zygosity groups. The implementation of biometric and psychometric models in LISREL can be achieved by approaches illustrated by McArdle and Goldsmith (1990), and Heath, Neale, Hewitt, Eaves, and Fulker (1989). The implementation of the saturated unconstrained model has

been illustrated by Neale, & Cardon (1992, p. 252).

The chi-square and the Tucker-Lewis (TLI, Tucker, & Lewis, 1973) index were used to select the most parsimoneous best fitting model.

Psychometric and biometric models are not nested, it is therefore not possible to use the chi-square difference test for models which belong to different classes. However, the psychometric one factor model is nested within the psychometric two factor model, and the biometric one factor model is nested within the biometric two factor model (all loadings on the second common factor are fixed at zero in the one factor model). Within each class, the chi-square difference test can therefore be used to test the one factor model against the two factor model.

The Tucker-Lewis index is defined as $TLI=(\chi_b^2/df_b-\chi_t^2/df_t)/(\chi_b^2/df_b-1)$. The TLI reflects the improvement in fit of a target model (subscript t) compared to a baseline model (subscript b). The TLI usually ranges from 0 to 1. Larger values imply a better fit. The TLI incorporates a penalty function for using more parameters, it may be poorer if additional parameters result in little improvement in χ^2. The TLI was preferred as an additional fit index because it is relatively independent of sample size (Marsh, Balla, & McDonald, 1988). The TLI was used to facilitate the comparison between the non-nested psychometric and biometric models. The chi-squares of the baseline versus the saturated, psychometric, biometric models were used to compute the TLI.

Results

For 1113 twins, both parents completed one CBCL/2-3 for each child. For 12 twin pairs only fathers' ratings were available, and for 247 twin pairs only mothers' ratings were available. For 4 twins, one child was rated by the mother and the other by the father, and for one twin pair only one child was rated. For 22 twin pairs there were at least two months between completion by the mother and the father. For these twins only the ratings of the fathers were used. For 32 twin pairs the father and the mother completed the questionnaire together.

Mean parental rating. The scales of the CBCL display considerable skewness and kurtosis. To perform accurate significance tests, weighted least squares estimation was used (Jöreskog & Sörbom, 1989 p. 22). For each twin pair, ratings on the 6 syndrome scales were available from both parents. There were therefore 24 variables in each zygosity group. With 24 variables, weighted least squares estimation would require a listwise sample size of at least 900 twin pairs in each zygosity group (Jöreskog & Sörbom, 1988 p. 3-32). In the phase of model selection, the mean of the parental ratings on each scale was used. When available, missing values for one parent were replaced by the rating of the other parent. This procedure reduced the number of variables to 12. For 12 variables the listwise sample size in the present study was large enough in both zygosity groups to use weighted least squares estimation. By computing the mean of the parental ratings, it is assumed that mothers and fathers observe the child's behavior in similar situations, and that they share a common understanding of the behavioral descriptions. Genetic analyses of maternal and paternal ratings of separate syndrome scales, supported this assumption (Van Den Oord, Verhulst, Boomsma, & Orlebeke, submitted).

Observed correlations between the 6 syndrome scales in the total twin sample are shown in Table 1. These correlations were computed with the mean parental score. The mean correlation was .40. Intercorrelations were somewhat higher among the Oppositional, Aggressive, and Overactive syndromes, and among the Withdrawn/Depressed

and Anxious syndromes.

Table 2 presents fit indices for the multivariate genetic models. Chi-squares of all models were significant at a level of .001. The lack of fit of the saturated model indicated that even when all correlations are simply decomposed in a genetic, shared environmental, and non-shared environmental contribution the model has to be rejected.

Tabel 1. Observed correlations for syndrome scales.

	1.	2.	3.	4.	5.	6.
1. Oppositional	1.000					
2. Withdrawn/depressed	.470	1.000				
3. Aggressive	.645	.402	1.000			
4. Anxious	.485	.490	.249	1.000		
5. Overactive	.573	.376	.539	.303	1.000	
6. Sleep Problems	.390	.247	.291	.301	.271	1.000

Biometric models fitted better than psychometric models. This suggested that genetic and environmental factors produce different patterns of problem behaviors. The chi-square difference test indicated that the two factor biometric model fitted significantly better than the biometric one factor model. The chi-square value of the biometric two factor model equaled almost the chi-square value of the saturated model. The TLI was even slightly higher for the biometric two factor model than for the saturated model. These findings suggested that the biometric two factor model accounted for the covariances between the syndromes almost as well as the model that simply decomposed all observed covariances between the syndromes in a genetic, shared environmental, and non-shared environmental contribution. The biometric two factor model was therefore preferred.

Tabel 2. Model fit indices for multivariate genetic models.

Model	common factors	df.	χ^2	TLI.
Baseline	none	138	1311.73	-
Saturated unconstrained		93	272.19	.774
Biometric	one	120	373.42	.752
Psychometric	one	130	605.77	.570
Biometric model*	two	105	289.35	.794
Psychometric	two	123	527.00	.614

Note. Weighted least squares estimation was used. * denotes preferred model.

It was also tested whether the biometric two factor model could be simplified by successively eliminating one genetic, one shared environmental, or one non-shared environmental factor. This was achieved by fixing all loadings on the second common factor at zero. Then the chi-square difference test was used to test the simplified model

against the full biometric two factor model.

The chi-square difference test indicated that eliminating the second genetic (χ^2_5=18.92), or the second non-shared environmental factor (χ^2_5=59.61) led to a significant ($p<.05$) decrease in fit. The decrease in fit was not significant when one shared environmental factor was eliminated (χ^2_5=4.12). All loadings on the second shared environmental factor in the biometric two factor model were low, which explained the non-significant decrease in fit. This simplification resulted in a biometric model with two common genetic, one common shared environmental, and two common non-shared environmental factors. The TLI of .804 also suggested that the simplified biometric two factor model should be preferred.

Parental ratings treated separately. To a large extent, parents in the present sample assessed the same behavior in their twins (Van Den Oord, Verhulst, Boomsma, & Orlebeke, submitted). In addition, there was a component which was unique to each rater. This unique component consisted of a random and a systematic part. The random part accounted for random errors of measurement such as scoring errors, misreading a question, and fluctuations of the parents' emotional state. The systematic part accounted for phenomena such as rater bias (the tendency of an individual rater to overstimate or underestimate scores consistently), or a somehow different understanding of the behavioral descriptions.

To obtain accurate parameter estimates it is neccessary to account for the unique part of each parents' rating. For example, random errors of measurement attenuate the observed covariances between the syndromes, and lead to underestimates of the effects of common factors. The preferred model was therefore specified in a way that accounted for the unique part of each parent's rating. This model was based on earlier findings concerning the properties of the maternal and paternal ratings in the present twin sample. Maternal and paternal ratings were viewed as two measurements of an underlying latent variable. This latent variable represented the behavior of the child as assessed by both parents. The latent variable was scaled by fixing the loading from the maternal and paternal rating on the latent variable at 1. Variances for the unique part of the maternal and paternal ratings were estimated for mothers and fathers separately. These variances were assumed to be equal for MZ and DZ twins. In addition, covariances between the unique part of each parent's rating of the first and second the twin were estimated. If genetic factors contribute to the unique part of each parent's rating, covariances between the unique parts of ratings of the first and second twin are larger for MZ than for DZ twins. These covariances were therefore estimated for MZ and DZ twins separately.

To fit the simplified biometric two factor model in a way that accounted for the unique part of each parents' rating, maternal and paternal ratings had to be included separately. For reasons mentioned above, it was not possible to use weighted least squares estimation anymore. Maximum likelihood estimation was used instead.

Parameters associated with the common part of the parental ratings are reported in Table 3. Input matrices were computed under pairwise deletion. For each zygosity group, the average of the pairwise sample sizes was used as the number of observations. The 32 twin pairs for whom the father and the mother completed the questionnaire together were discarded from these analyses.

Table 3 shows that for the common factors genetic effects were larger than shared environmental effects, and shared environmental effects were larger than non shared environmental effects. On the average the common factors explained 72% of the variance of the syndromes. Genetic influences accounted on the avarage for 44% of this common

Tabel 3. Parameter esimates associated with the common part of the parental ratings.

	first common factor			*second common factor*			*unique factors*		
Syndrome	A	C	E	A	C	E	A	C	E
Oppositional	2.016	3.088	1.715	2.618	-	.096	.000	.000	2.510
Withdr./Depr.	.358	.381	.152	.611	-	.266	.519	.000	.536
Aggressive	1.606	1.243	.722	.498	-	-.267	.000	.000	.675
Anxious	-	1.096	-	1.709	-	1.694	.000	.000	.000
Overactive	.597	.646	.940	.783	-	-.192	.000	.000	1.035
Sleep Pr.	-.109	.981	.315	.318	-	-.032	.869	.000	1.107

Note. A is additive genetic factor, C is shared environmental factor, E is non-shared environmental factor. Maximum likelihood estimation was used. - denotes parameter fixed at zero.

variance, shared environmental influences for 38%, non-shared environmental influences for 19%. These findings indicated that shared environmental influences were almost as important as genetic factors, and non-shared environmental influences were least important for causing covariances between syndromes.

Non-shared environmental influences were most specific in their effects. There were no effects of the unique shared environmental factors. The different syndromes were identical as shared environmental influences were concerned.

Estimates of parameters associated with the unique part of the parental ratings are reported in Table 4.

Tabel 4. Parameter esimates associated with the unique part of the parental ratings.

	maternal ratings		*paternal ratings*	
Syndrome	variance	covariance	variance	covariance
Oppositional	13.046	9.168/6.797	9.531	3.870/3.067
Withdr./Depr.	1.183	.758/.343	.921	.626/.339
Aggressive	2.395	1.335/.779	2.036	1.064/.512
Anxious	3.129	1.930/.983	2.777	.976/.830
Overactive	1.433	.731/.115	1.521	.368/.236
Sleep Pr.	1.406	.976/.321	1.178	.762/.390

Note. Maximum likelihood estimation was used. Numbers before slash apply to MZ twins, numbers after slash apply to DZ twins.

Results from Table 3 and Table 4 suggested that on the average 32% of the variance of the maternal and paternal ratings consisted of unique variance. The covariances between the unique part of each parent's rating of the first and second the twin were somewhat higher for MZ than for DZ twins. This suggested that besides error of

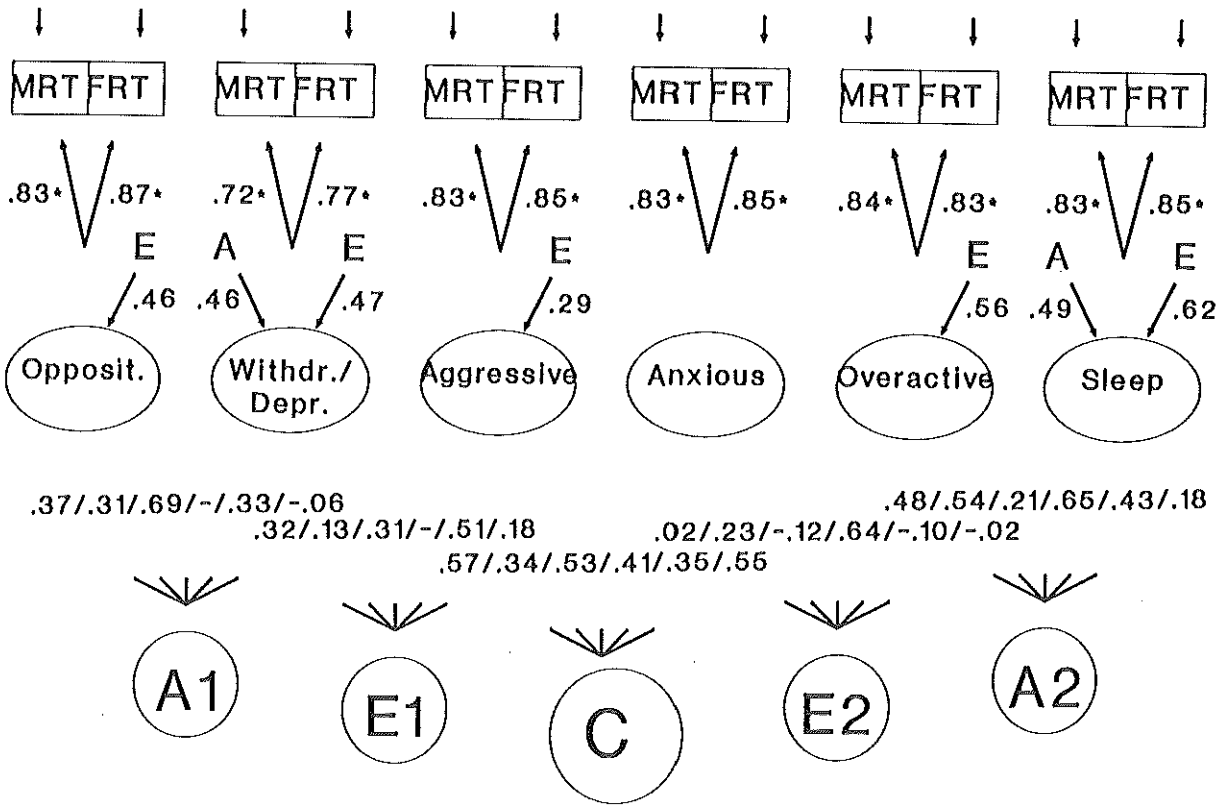


Figure 1. Preferred multivariate genetic model. The completely standardized solution is presented. A is additive genetic factor, C is shared environmental factor, E is non-shared environmental factor. MRT is maternal rating, FRT is paternal rating. Unique factors for which a zero effect was estimated are ommitted. * denotes a rescaled fixed parameter. Estimates of the variances and covariances for the unique part of the parental ratings are not presented.

measurement and rater bias, genetic influences also contribute to the unique part of each parent's rating.

Figure 1 depicts the completely standardized solution. In this solution, all latent and observed variables are scaled the have variances equal to one. There were substantial loadings from all syndromes on the common shared environmental factor. This suggested that a pattern of similar scores on all syndromes is indicative of a shared environmental cause.

The Oppositional, Withdrawn/depressed, and Overactive syndromes loaded on both common genetic factors. The Aggressive syndrome loaded especially on the first common genetic factor. The Anxious syndrome loaded especially on the second common genetic factor. This indicated that with respect to the common genetic factors only the Aggressive and Anxious syndromes were distinct syndromes. The first common genetic factor produced a clustering of the Oppositional, Withdrawn/Depressed, and Overactive syndromes with the Aggressive syndrome. The second common genetic factor produced a clustering of the Oppositional, Withdrawn/Depressed, and Overactive syndromes with the Anxious syndrome.

Especially the Oppostional, Aggressive, and Overactive syndromes loaded on the first common non-shared environmental factor. This indicated that a so-called externalizing pattern of problem behaviors is most suggestive of the first common non-shared environmental factor. An internalizing pattern of problem behaviors appeared to be more suggestive of the second common non-shared environmental factor.

Discussion

Multivariate genetic models were fitted to study patterns of problem behaviors in three-year-old twins. A biometric model with two common genetic, one common shared environmental, and two common non-shared environmental factors accounted almost as well as the saturated unconstrained model for the genetic and environmental covariances. The common non-shared environmental factors produced externalizing/internalizing patterns of problem behaviors. One common genetic factor produced a clustering of the Oppositional, Withdrawn/Depressed, and Overactive syndromes with the Aggressive syndrome. The other common genetic factor produced a clustering of the Oppositional, Withdrawn/Depressed, and Overactive syndromes with the Anxious syndrome. A pattern of similar scores on all dimensions of problem behavior was most suggestive of the common shared environmental factor.

For children with multiple problems, patterns of problem behaviors could be used to distinguish children with high test scores into groups with high genetic or high environmental scores. In the twin sample, most efficiently this could be done with factor estimation procedures (Boomsma, Molenaar, and Orlebeke, 1990). In addition, standard errors of the estimate could be obtained. Such a differential diagnosis at the level of etiology might be useful for several reasons. For instance, groups of children could be compared with respect to response to treatment. Knowledge about a possible differential response, could eventually be used to optimalize clinical interventions.

Exploratory higher order factor analyses of Dutch syndrome scales yielded an Externalizing grouping comprising the Oppositional, Aggressive, Overactive syndromes and an Internalizing grouping comprising the Anxious, Withdrawn/Depressed syndromes (Koot, Van Den Oord, Verhulst, & Boomsma, submitted). However, confirmatory factor analyses indicated that a two factor model did not fit and suggested that Exter-

nalizing/Internalizing could not be interpreted as two higher order factors. Higher order factor analyses implicitly assume that common genetic and environmental factors influence the syndromes via the higher order factors, and that common genetic and environmental factors display similar patterns. Results from the present study showed that genetic and environmental influences influence syndromes directly, and that they produce different patterns. This could explain the failure of the two factor model in the higher order factor analyses. Only the common non-shared environmental factors showed internalizing or externalizing patterns of problem behaviors. With respect to the common genetic and shared environmental factors there was no distinction between internalizing and externalizing behaviors. This may also explain the difficulties in obtaining a sharp distinction between Externalizing/Internalizing behaviors in the higher order factor analyses of Dutch syndromes. The mean factor loading in three independent samples was for the withdrawn syndrome .461 on the "Internalizing" factor and .294 on the "Externalizing" factor. This resembled American findings. Achenbach (1992) found for the American Withdrawn scale equal loadings on both the "Internalizing" and "Externalizing" factors.

CBCL studies consistently showed larger shared environmental influences for total problem scores than for any of the separate scales (Edelbrock, Rende, & Plomin, 1992; Van Den Oord, Boomsma, & Verhulst, submitted; Van Den Oord, Verhulst, Boomsma, & Orlebeke, submitted). In addition, Plomin, DeFries, and Fulker (1988, pp. 183-184) did not study separate CBCL scales but they did find evidence for substantial shared environmental influences on the CBCL total problem score. The present study indicated that shared environmental influences were especially important for causing covariances between syndromes. This could explain the larger shared environmental influences on CBCL total problem scores than for separate syndrome scales. The variance of the total problem score consists of the variance of separate syndrome scales plus two times all the covariances between the scales. The relatively large shared environmental influences on covariances between different syndromes affects the univariate genetic analyses of the total problem score, but not of the separate syndromes.

Fit indices indicated that even the saturated model should be rejected. An earlier paper showed that there were small differences between the distribution of problem scores in MZ versus DZ twins (Van Den Oord, Koot, Boomsma, Verhulst, & Orlebeke, submitted), and genetic analyses yielded some evidence for sibling interaction and possibly for non-additive genetic variance (Van Den Oord, Verhulst, Boomsma, & Orlebeke, submitted). The relatively large sample size and the use of a multivariate test instead of univariate tests may have increased the power to detect these kind of effects, and contributed to failure of the saturated model.

A genetic study of separate syndrome scales (Van Den Oord, Verhulst, Boomsma, & Orlebeke, submitted) suggested that on the average genetic, shared environmental, and non-shared environmental influences accounted for 65%, 12%, and 21% of the variance. In the present study, these percentages were respectively 41%, 22%, and 37%. Two factors may have contributed to the different percentages found in the present multivariate study compared to the percentages found in the univariate study. A low power to detect modest shared environmental effects in univariate genetic analyses (Martin, Eaves, Kearsey, & Davies, 1978), may have resulted in non-significant shared environmental effects for some scales in the univariate genetic analyses. This could explain the lower percentage shared environmental influences found in the univariate study. In the univariate genetic analysis, allowing sibling contrast effects for the Overactive syndrome

yielded a superior fit and larger genetic effects. In the present study, sibling effects were not allowed for the Overactive syndrome, because no acceptable parameter estimate was obtained. This could explain the smaller genetic effect found in the present study for the Overactive syndrome, but it is unlikely that it explains the smaller heritabilities found for the other syndromes too. For each scale, the mean of the 2 correlations between the maternal rating of one twin and the paternal rating of the other was computed. For the six scales, the average of these mean twin correlations was for MZ twins .50, for DZ twins .20. These correlations suggested that the genetic influences were somewhat underestimated in the present study, and a heritability that was more in agreement with the genetic analyses of separate syndrome scales.

9
Discussion

Introduction

In this chapter, results from the genetic analyses are summarized and discussed. Attention is paid to the interpretation of the findings, and recommendations for future research are made. Finally, issues concerning use and misuse of genetic findings are addressed.

Conclusions

The primary aim of this dissertation was to study genetic and environmental influences on problem behaviors in young children and adolescents. Quantitative genetic analyses were performed on a sample of 11-15-year-old international adoptees (111 pairs of biological siblings, 221 pairs non-biological siblings, 94 singletons), and on a sample of 3-year-old twins (407 pairs of MZ twins, and 874 pairs of DZ twins).

In the sample of international adoptees, genetic, shared environmental, and non-shared environmental influences accounted on average for 29%, 18%, and 53% of the variance of the syndrome scales. Genetic influences were much more important for externalizing than for internalizing behaviors. For internalizing behaviors, genetic, shared environmental, and non-shared environmental influences accounted for 2%, 20%, and 78% of the variance, for externalizing behaviors these percentages were 63%, 15%, and 23%. Significant sex differences in genetic and environmental influences on problem behaviors were found. Genetic influences were largest for externalizing behaviors in boys. For a number of problem behaviors, shared environmental influences were somewhat larger for girls than for boys.

In the twin sample, genetic, shared environmental, and non-shared environmental influences accounted on average for 65%, 12%, and 21% of the variance. There were no clear differences between genetic and environmental influences on internalizing or externalizing behaviors. Sex differences in genetic and environmental influences on problem behaviors were small.

Discussion

In chapter 2 it was noted that small sample sizes, the use of instruments for which validity and reliability were not clearly demonstrated or that assessed such a broad array of problem behaviors that it became unclear what was measured, and the use of possibly inappropriate genetic models made it difficult to draw firm conclusions concerning genetic and environmental influences on problem behaviors in children. In the present study sample sizes were clearly larger compared to most previous genetic studies of problem behaviors in children, item analyses were performed to establish the properties of the assessment instrument and to derive narrowly defined syndromes, and the validity of the genetic models was tested and, when neccessary, adaptations were made to obtain more valid estimates of genetic and environmental effects. Nevertheless, a number of points that are important or that may have affected the results should be mentioned.

Genetic influences were substantial for most problem behaviors. Only for

internalizing behaviors in the sample of international adoptees, no evidence was found for genetic influences. This was in contrast to the genetic studies of problem behaviors in 4-18-year-olds reported in the literature review. Compared to the findings in the adoption sample and the genetic studies of problem behaviors in children aged 4-18 years as were reported in the literature, heritabilities were larger in the twin sample. Partially this can be attributed to the use of a more accurate model in the twin sample. Because both parents rated each twin, a statistical model could be fitted that decomposed the total variance into trait variance and variance that was associated with properties of the rater or ratings (see chapter 7). This latter part comprises variance due to unreliability and rater bias (a tendency of an individual rater to over- or underestimate scores consistently). When not accounted for, unreliability of the assessment instrument spuriously inflates estimates of the non-shared environment and rater bias spuriously inflates estimates of the shared environment. As a consequence the relative importance of genetic influences (the heritability) is underestimated. However, part of the larger heritabilities in the twin sample may also reflect that for younger children genetic influences are more important than for older children.

In both the adoption and the twin sample, environmental influences which are shared by children within the same family were smaller than either genetic or non-shared environmental influences. This finding agrees with genetic studies in other areas such as personality and cognition (Plomin, & Daniels, 1987).

Sex differences were substantial in the adoption sample, but small in the sample of 3-year-old twins. This finding suggests that sex differences in genetic and environmental influences are larger in older children. It is also in agreement with large scale studies which demonstrated that sex differences in prevalence rates and levels of problem behaviors are less consistently found in younger children (Campbell, 1989).

Generalizability to child psychiatric conditions. To generalize findings from the present study to child psychiatric conditions, it has to be assumed that these conditions represent extremes on the same continuum that describes variation within the normal range. Although this assumption may very well apply to the commoner varieties of emotional and conduct disturbances (Rutter et al., 1990b; Plomin, 1990), there may be qualitative as well as quantitative distinctions too (Rutter et al., 1990b). If this assumption is incorrect, findings from the present study cannot be extrapolated to psychiatric conditions. For instance, if clinical depressions are affected by other genes or other environmental factors than "mood" differences between children in the general population, genetic and environmental etiologies may be quite different for clinical and non-clinical populations.

A related issue is that genetic and environmental etiologies that completely explain a disorder for a few individuals, account for a negligible amount of variance in the population as a whole and could thus remain undetected in our analyses of total variations in the population (Plomin, Rende, & Rutter, in press).

Assortative mating. A number of studies showed low-level positive correlations (average .10 to .15) between spouses for several personality traits (Price, & Vandenberg, 1981). This spouse resemblance in personality traits appears to arise primarily from initial mate selection rather than the effect of living together (Buss, 1984; Mascie-Taylor, 1989; Phillips, Fulker, Carey, & Nagoshi, 1987; Price, & Vandenburg, 1981). The tendency of mates to select partners like themselves is called positive assortative mating. For psychopathology, the presence of mental disorder in one parent also appears to be associated with an increased risk of the disorder in the other (Hagnel, & Kreitman, 1974; Merikangas, Weissman, Prusoff, & John, 1988; Quinton, & Rutter, 1984, pp. 106-107).

Assortative mating may constitute part of the explanation (Merikangas, Weissman, Prusoff, & John, 1988), but evidence also suggests some sort of 'contagion' effect (Rutter, & Cox, 1985). Thus, the co-occurrance of disorders in both spouses may reflect maladaptive features of marital interactions or a family burden imposed by noxious events.

Strictly speaking, assortative mating is only of importance to the extent that the genotypes of the parents were correlated when they were the same age as their children are now. In this case, positive assortative mating tends to increase the similarity of DZ twins relative to MZ twins (Neale, & Cardon, 1992, p. 19). Consequently, positive assortative mating will artificially inflate estimates of the shared environmental, whereas the genetic component will tend to be biased downwards. Thus, if positive assortative mating had occurred, this would have led to underestimates of the heritability.

It could also be that mates select partners on the basis of other, instead of similar, traits. For instance, depressive people could prefer antisocial partners. This kind of assortative mating would result in associations between these different traits in spouses as well as in their children (Thompson, 1966). In a multivariate genetic analysis (see chapter 8) these associations might spuriously be considered as patterns of problem behaviors that arise from the fact that different problem behaviors are variable expressions of the same genetic or environmental cause.

Variances and prevalence rates. Under assumption that quantitative test scores are liabilities or 'risks' to psychiatric conditions, differences in variances may have implications for prevalence rates. In the sample of international adoptees, genetic influences accounted for the larger variance of externalizing behaviors in boys. Figure 1 illustrates how this larger variance, caused by genetic influences, might contribute to the higher prevalence found for externalizing behaviors in boys (Verhulst, & Koot, 1992).

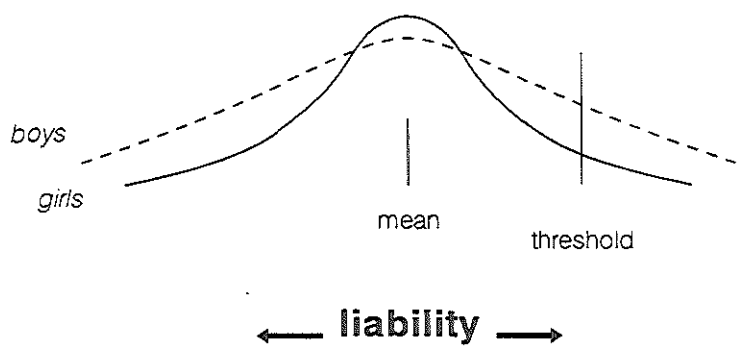


Figure 1. A model for relating liabilities to prevalence rates.

The model in Figure 1 assumes that when liabilities exceed a threshold, the disorder may be classified as present. The prevalence rate equals the surface beyond the threshold and under the right tale of the distribution. Because of the larger variance, this surface is

larger for boys. It should be mentioned that in this model, the prevalence rates also depend on the mean of the distribution and the value of the threshold. Both a higher mean liability and a lower threshold for boys, could therefore also contribute to the higher prevalence rate found for externalizing behaviors in boys.

Zygosity determination and parental expectation. For the majority of same-sex twins, zygosity was determined through a questionnaire completed by parents when their twins were about 2 years old. To establish the reliability of the questionnaire, blood test results were compared with the zygosity information on the questionnaire. From the 327 same sex twin pairs for whom both blood test and questionnaire results were available, 61 twin pairs were misclassified (19%). Fourteen DZ twin pairs were misclassified as MZ, 47 MZ twin pairs were misclassified as DZ. It is likely that parents who were uncertain about their twins' zygosity were more likely to consent with the blood test. Perhaps, this percentage is therefore better viewed as a lower bound of the reliability of the questionnaire.

It is useful to note how misclassifications could have affected the results in the present study. When genetic factors influence a trait, misclassifying MZ twins as DZ twins will overestimate the "DZ" twin similarity. Misclassifying DZ as MZ twins will underestimate the "MZ" twin resemblance. Misclassifications therefore attenuate the difference between the MZ and DZ twin correlations. Consequently, genetic influences are underestimated and shared environmental influences are overestimated.

An important assumption in twin research is that the environments of MZ twins are no more similar than the environments of DZ twins. For instance, MZ twins could be treated more alike by other people than DZ twins. A related issue is that parents might expect their MZ twins to develop along similar lines. This expectation may become self-fulfilling or spuriously inflate the parental ratings of their twin's resemblance. The greater similarity of MZ twins would then result in overestimates of genetic influences, and underestimates of the shared environmental influences.

By comparing twin correlations in correctly classified and misclassified twins, this assumption can be tested. If twin correlations differ for correctly and incorrectly classified twins, this may reflect an expectancy effect. Goodman, and Stevenson (1989b) argue that MZ twins misclassified as DZ twins are more useful for this test than DZ twins misclassified as MZ twins. Some DZ twin pairs will be particularly alike because they share, for instance by chance, an unusual high proportion of the relevant genes. If these unusual similar DZ twin pairs are particularly likely to be misclassified as MZ, than misclassified DZ twins will be more alike than correctly classified DZ twins. It should be mentioned that it is also possible that MZ twins are thought to be DZ because of non-genetic differences. For instance, the risk of birth injury may be different for the first and second twin (Bulmer, 1970, pp. 62-64). Such non-genetic factors may make MZ twins less alike, and could increase the chance of misclassification.

A number of studies have compared twin correlations of correctly and incorrectly classified MZ twins. Cohen, Dibble, & Crawe (1977) found no differences between a group of misclassified MZ twins (25 pairs) and correctly classified MZ twins (130 pairs) for mother and father ratings of personality characteristics. Scarr and Carter-Salzman (1979) comparing 84 pairs correctly and 19 pairs incorrectly classified twin pairs, found no differences for measures of extraversion and self esteem. In contrast, Goodman, and Stevenson (1989b) found for an (hyper)activity scale somewhat lower twin correlations for mother, father, and teachers ratings of MZ pairs mistakenly thought by their parents to be DZ. Twin correlations for mothers', fathers', and teachers' ratings of hyperactivity for

MZ twins who were correctly identified as MZ twins by both parents were .71(69 pairs), .56(57 pairs), .69(66 pairs). For MZ pairs mistakenly thought by parents to be non-identical twin correlations for mothers', fathers', and teachers' ratings of hyperactivity were .64(23 pairs), .42(17 pairs), .47(23 pairs).

In the present study a similar comparison could be made. Blood test results indicated that 47 MZ twins were, according to the parental response on the questionnaire, mistakenly thought of as DZ twins. It should be noted that the questionnaire was completed when the twins were 2 years old, while the ratings of problem behaviors were obtained when the twins were 3 years old. The parental perception of the twins' zygosity may have changed during this period. Parental reports of the zygosity of their twin were also obtained when the twins were 1 year old. It appeared that 17% of the 47 parents had changed their minds concerning their twin's zygosity, between the time their twins were 1 and 2 years old. Similar changes could also have occurred during the time the twins were 2 years old, and the time the ratings of problem behaviors were obtained. At the time the parents rated problem behaviors in their twins, blood test results were not available yet and consequently unknown to the parents.

To compute the twin correlations we used for each scale the mean of the 2 correlations between the maternal rating of one twin and the paternal rating of the other. This procedure is consistent with the genetic analyses in the twin sample, in which basically the same correlation was used to obtain heritability estimates. The average twin correlation for the seven syndromes was .44 in the group of MZ twins misclassified as DZ. For the 138 MZ twins who were, according to the blood test, correctly classified as MZ twins the average twin correlation was .54. Twin correlations tended to be somewhat higher for the correctly classified MZ twins. This could be the result of an expectancy effect. However, the difference was small and could also be explained by sample fluctuations.

It could also be that both the parents of MZ and DZ twins regard their children as more similar just because they are twins. This would result in overestimates of the shared environmental influences, because it increases the DZ as well as the MZ twin correlation.

Interpretation of the results

Heritability and the nature-nurture issue. High heritabilities do have important implications for the nature-nurture issue. For instance, a heritability of .65 as was found in the twin sample means that 65% of differences between childrens' problem behaviors are innate. This suggests that children with behavior problems are likely to show an innate vulnerability. It should be noted that, as stressed by many authors, a high heritability does not mean that the behavior of concern is unchangeable (e.g. Plomin, & Daniels, 1986; Rutter, 1991; Vandenberg, & Crowe, 1989). The finding of genetic effects implies hereditary propensities, not predestination (Plomin & Daniels, 1986). Furthermore, the heritablity is an index for average differences among individuals in a population. Moderate heritability in a population could therefore mask total environmental etiology for some individuals and total genetic etiology for others.

Heritability and familial resemblance. Genetic analyses of separate syndrome scales indicated that genetic influences accounted on average for 65% of the differences between three-year-old twins. A heritability of similar magnitude was obtained for externalizing behaviors in the sample of international adoptees. How should this finding be interpreted? Does it mean that children from one family should be highly alike, or that children of

parents with psychological complaints are at high risk of developing behavior problems?

Heritability is often associated with resemblance between relatives. Indeed, when hereditary factors influence a trait, siblings become alike because they share a proportion of their genetic information, and children may resemble their parents from whom they received their genes. However, some points need to be kept in mind when inferring resemblance between relatives from heritability estimates.

As shown in chapter 2, the resemblance between full siblings can be expressed as (assuming a model with additive genetic effects, shared environmental effects, and non-shared environmental effects): $r=.5h^2+c^2$ (in which r is the observed correlation between siblings, h^2 the heritability, and c^2 the proportion of shared environmental variance). The formula shows that the genetic contribution to the observed sibling correlation is half the heritability. A heritability of .65 therefore implies a sibling correlation of .325. In the social sciences, a correlation of this size may be regarded as "medium" (Cohen, 1988, pp. 79-81).

The formula for sibling resemblance ($r=.5h^2+c^2$) also illustrates that environmental influences which are shared by children within the same family, are potentially more powerful for creating sibling resemblance than genetic influences. This is because these environmental influences are identical for siblings. However, results from the present study indicated that shared environmental influences were small. For sibling resemblance in problem behaviors, genetic influences seem therefore the most important source.

Hereditary factors may also influence the resemblance between parents and their children. However, this is only the case when the same genes influence behavior in both parents and children. For this reason, the behavior of children is perhaps better compared with the behavior of their parents when they were the same age as their children are now. When the expression of genes is not age dependent, the genetic correlation between parents and children equals .50, and the parent-child correlation equals half the heritability. With a heritability of .65 this correlation may, again according to Cohen's criteria, be considered "medium".

Heritabilities can also be used to predict childrens' scores from the parental scores, through regression analysis (Falconer, 1989, p. 167). When information of one parent is known the reduction in prediction error equals the square of half the heritability. For example, with a heritability of .65 this corresponds with a reduction of prediction error variance of 10%. In the social sciences (Cohen, 1988, pp. 79-81) this may be regarded as a medium effect size. When information of both parents is available, the children's scores can be regressed on the mid-parental value. In this case the reduction in prediction error variance equals the square of the heritability (Falconer, 1990, p. 153). In our example this would correspond with a reduction of prediction error variance of 43% which in the social sciences may be regarded as a large effect size (Cohen, 1988, pp. 79-81).

The example presented above illustrates that a high heritability, as was found in the present study, results in "medium" resemblance between separate members of a family. Under assumption that the expression of genes is not age dependent, psychopathology in parents is a predictor of problem behaviors in children. With high heritabilities, an accurate prediction can be obtained when psychopathology is assessed in both parents. This suggests that maladjustment in both parents, may from a genetic point of view imply that their children are at considerable risk for developing behavior problems.

Recommendations for future research

Concerning results from the genetic analyses on the sample of 11-15-year-old adoptees and the 3-year-old twins, there were some marked differences. Especially for internalizing behaviors, heritabilities were smaller in the adoption sample. Furthermore, sex differences were substantial in the adoption sample, but small in the twin sample. It would be interesting to know whether these differences represented true age effects, or reflected sampling error or sample differences. A genetic study, for instance in a Dutch adolescent twin sample, would therefore be useful to obtain more information concerning the cause of these differences.

To generalize findings from genetic studies in non-clinical samples to child psychiatric conditions, it has to be assumed that these conditions represent extremes on the same continuum that describes variation within the normal range. This issue seems to be too important to be merely a supposition, and would require an empirical test (Plomin, in press; Rutter, 1988). DeFries, and Fulker (1985, 1988) have shown how this problem may be approached from a quantitative genetic perspective. A drawback of their method is that very large numbers of twin pairs would be required to test this hyphothesis that disturbed children represent the lower tale of a normal of individual differences (DeFries, & Fulker, 1988). However, with the establishment of twin registers these numbers may become feasible in the future.

Assortative mating was mentioned as a factor that may have affected the results in the twin sample. In order to determine the precise effect, it would be useful to obtain information about psychopathology in the parents of the twins.

In chapter 8 it was suggested that patterns of problem behaviors could be used to distinguish children with high test scores into groups with high genetic or high environmental scores. Such a differential diagnosis at the level of etiology might be useful for several reasons. For instance, groups of children could be compared with respect to response to treatment in order to optimalize clinical interventions. In addition, in a longitudinal study, the stability of problem behavior in the different groups could be determined. Knowledge about differential stability could be used to distinquish between groups of children who require immediate interventions versus children whose problem seems to be temporary.

Use and misuse

Even today genetic issues remain highly sensitive (Vandenberg, & Crowe, 1989). A fatalistic view that we could do nothing about genetic effects, and a distaste arising from the misuse of genetics in support of racist and eugenic policies may explain some of the sensitivity of the subject (Rutter, 1991).

Rose, Kamin, and Lewontin (1984) discuss a large number of examples of past and recent abuse of genetic findings and arguments. Interestingly, several examples concern child psychiatric conditions. The authors offer an elaborate discussion (pp. 178-188) about children who show problem behavior in the class room. For the United States, children who were overactive, had concentration problems, and interrupted the teacher, suddenly became sick during the 1960s. These problems were defined as biological and medical in nature. Terms like "minimal brain damage" and "minimal brain dysfunction" came into common usage. The proposed remedy was to treat the offending children with drugs. Within a couple of years many hundreds of thousands American schoolchildren labeled as MBD, hyperactive, or learning disabled were receiving medication. Rose,

Kamin, and Lewontin (1984) view the rise of genetic research on these kind of problems as an attempt to find even more "proof" of a biological defect. Demonstrating a genetic basis for this disorder would locate the problem within the child and further justify treating the "defect".

The point that Rose, Kamin, and Lewontin (1984) make is that genetic arguments can be misused to serve as a rationale to understand and to cope with deviance. Reducing disorders to biological defects in the child may be a justification to control and pacify unruly children with medical treatment or special education.

We clearly recognize that locating a problem in a child may benefit others for a variety of reasons such as shortcomings in teaching or clinical skills, to mask problems of parents themselves, or simply as a consequence of personal dislike. We also recognize that genetic research may provide a rationale. However, we do not see any justification for this. Behavior genetics studies individual differences. Labels such as inequal or deviant are evaluations of these differences that do not follow from quantitative genetic theory, but are made by people who might benefit from doing so. On the contrary, by accounting for individual differences between children, behavior genetics advocates a greater recognition of and respect for individuality. An acceptance of differences between children is much less in agreement with labels such as deviant, than a denial of individuality is.

Problem behaviors in children not only exist for parents and teachers, but also for the children themselves. Part of understanding and dealing with these problems lies in recognizing innate differences (Plomin, & Daniels, 1986). Not all children who experience noxious situations are bound to become problem children, and other children simply seem to more inclined to show behavior problems. Explaining behavior problems entirely from an environmental perspective seems not only to be in disagreement with reality, it could also harm those that are involved. Parents often feel that they did something wrong. Attributing causes of problem behaviors entirely to the child's environment, from which parents are such an important part, might make them feel even more quilty. It could also harm the troubled children and youth themselves, by denying that some children require special attention and have specific needs. We hope that a broader recognition of innate differences in liabilities might help to prevent problem behaviors by making caretakers more aware of the specific needs of some children, and, when neccessary, helps to influence the child's behavior in order to achieve more satisfactory levels of functioning.

References

Achenbach, T.M.(1966). The classification of children's psychiatric symptoms: A factor-analytic study. Psychological Monographs, 80 (Whole No. 615).

Achenbach, T.M.(1978). The Child Behavior Profile: I. Boys aged 6-11. Journal of Consulting and Clinical Psychology, 46, 478-488.

Achenbach, T.M.(1988). Integrating assessment and taxonomy. In Rutter, R., Tuma, A.H. and Lann, I.S.(eds), Assessment and Diagnosis in Child Psychopathology. London: Fulton.

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. 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.

Achenbach, T.M., & Edelbrock, C.S.(1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four to sixteen. Monographs of the Society for Research in Child Development, 46 (Serial number 188).

Achenbach, T.M., & Edelbrock, C.S.(1983). Manual for the Child Behavior Checklist and Revised Child Profile. Burlington, Vt.: University of Vermont, Department of Psychiatry.

Achenbach, T. M., Edelbrock, C., & Howell, C. T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3-year old children. Journal of Abnormal Child Psychology, 15, 629-650.

Achenbach, T.M., McConaughy, S.H., & Howell, C.T.(1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. Psychological Bulletin, 101, 213-332.

American Psychiatric Association(1987). Diagnostic and Statistical Manual of Mental Disorders,(third edition revised). Washington D.C.: American Psychiatric Association.

Anderson, J.C., & Gerbing, D.W.(1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. Psychometrica, 49, 155-173.

Bakwin, H.(1971). Enuresis in twins. American Journal of Disease in Childhood, 121, 222-225.

Bates, J.E., Freeland, C.A., & Lounsbury, M.(1979). Measurement of infant difficultness. Child Development, 50, 794-803.

Behar, L. B., & Stringfield, S.(1974). A behavior rating scale for the preschool child. Developmental Psychology, 10, 601-610.

Bentler, P.M., & Bonnet, D.G.(1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Berg, I.(1976). School phobia in the children of agoraphobic women. British Journal of Psychiatry, 128, 86-90.

Biederman, J., Munir, K., Knee, D., Habelow, W., Armentano, M., Autor, S., Hoge, S.K., & Waternaux, C.(1986). A family study of patients with attention deficit disorder and normal controls. Journal of Psychiatric Research, 20, 263-274.

Bohman, M.(1971). A comparative study of adopted children, foster children and children in their biological environment born after undesired pregnancies.Acta Paediatrica Scandinavica, Supplement 221.

Bohman, M.(1972). The study of adopted children, their background, environment and adjustment. Acta Paediatrica Scandinavica, 61, 90-97.

Bohman, M.(1982). The interaction of heredity and childhood environment: some adoption studies. Journal of Child Psychology and Psychiatry, 22, 195-200.

Bohman, M., Cloninger, R., Sigvardsson, and von Knorring, A.(1982). Predisposition to petty criminality in swedish adoptees.Archives of General Psychiatry, 39, 1233-1240.

Bohman, M., Sigvardsson, S.(1980). A prospective, longitudinal study of children registered for

adoption. Acta Psychiatrica Scandinavica, 61, 339-355.
Bollen, K.A.(1989). Structural Equations with Latent Variables. New York: Wiley.
Boomsma, D.I., & Molenaar, P.C.M.(1986). Using LISREL to analyze genetic and environment covariance structures. Behavioral Genetics, 16, 237-250.
Boomsma, D.I., Molenaar, P.C.M., & J.F. Orlebeke (1990). Estimation of individual genetic and environmental factor scores. Genetic Epidemiology, 7, 83-91.
Bulmer, M.G.(1970). The Biology of Twinning in Man. Oxford: Clarendon.
Buss, D.M.(1984). Marital assortment for personality dispositions: Assessment with three different data sources. Behavior genetics, 14, 111-124.
Buss, A.H., & Plomin, R.(1975). A Temperament Theory of Personality Development. New York: Wiley-Interscience.
Buss, A.H., & Plomin, R. (1984). Temperament: Early Developing Personality Traits. Hillsdale, NJ: Erlbaum.
Cadoret, R.J.(1978). Psychopathology in adopted away offspring of biological mothers with antisocial behavior. Archives of General Psychiatry, 35, 176-189.
Cadoret, R.J.(1986). Adoption studies: Historical and methodological critique. Psychiatric Developments, 1, 45-64.
Cadoret, R.J., & Cain, C.(1980). Sex differences in predictors of antisocial behaviors in adoptees. Archives of General Psychiatry, 37, 1171-1175.
Cadoret, R.J., Cain, C.A., & Crowe, R.R.(1983). Evidence for gene-environment interaction in the development of adolescent antisocial behavior. Behavior Genetics, 13, 301-311.
Cadoret, R., Cunningham, L., Loftus, R., & Edwards, J.(1975). Studies of adoptees from psychia trically disturbed biological parents. II. Temperament, hyperactive, antisocial and developmental variables. Journal of Pediatrics, 87, 301-306.
Cohen, D.J., Dibble, E., & Grawe, J.M. (1977). Fathers' and Mothers' perceptions of children's personality. Archives of General Psychiatry, 34, 480-487.
Cohen, J.(1988). Statistical Power Analysis for the Behavioral Sciences. Hillsdale: Erlbaum.
Campbell, S.B.(1989). Developmental perspectives. In T.H. Ollendick and M. Hersen(eds), Handbook of Child Psychopathology,(5-29). New York: Plenum.
Campbell, S. B.(1990). Behavior Problems in Preschool Children: Clinical and Developmental Issues. New York: The Guilford Press.
Cantwell, D.P.(1972). Psychiatric illness in the families of hyperactive children. Archives of General Psychiatry, 27, 414-417.
Cantwell, D.P.(1975). Genetic studies of hyperactive children: Psychiatric illness in biologic and adopting parents. In R.R. Fieve, D. Rosenthal, and H. Brill(eds), Genetic Research in Psychiatry, (pp. 273-280). Baltimore: John Hopkins University Press.
Carey, G.(1986). Sibling imitation and contrast effects. Behavioral Genetics, 16, 319-341.
Carey, G.(1992). Twin imitation for antisocial behavior: implications for genetic and family environment research. Journal of Abnormal Psychology, 101, 18-25.
Caron, C., & Rutter, M.(1991). Journal of Child Psychology and Psychiatry, 32, 1063-1081.
Centraal Bureau voor de Statistiek.(1989). Standaard Onderwijsindeling SOI-1978. Voorburg/Heerlen: Centraal Bureau voor Statistiek.
Cheek, J.M., and Buss, A.H.(1981). Shyness and sociability. Journal of Personality and Social Psychology, 41, 330-339.
Cicchetti, D., & Toth, S. L.(1991). Rochester Symposium on Developmental Psychopathology, Vol. II: Internalizing and externalizing expressions of dysfunction. Hillsdale, NJ: Erlbaum.
Cohen, D.J., Dibble, E., & Grawe, J.M.(1977). Fathers' and Mothers' perceptions of children's personality. Archives of General Psychiatry, 34, 480-487.
Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart, and Winston.

Crowe, R.R.(1974). An adoption study of antisocial personality. Archives of General Psychiatry, 31, 785-791.

Cunningham, L., Cadoret, R.J., Loftus, R., & Edwards, J.E.(1975). Studies of adoptees from psychiatrically disturbed biological parents: II Temperament, huperactive, antisocial and developmental variables. Journal of Pediatrics, 87, 301-306.

Dahlstrom, W.G., & Welsh, G.S.(1960). An MMPI Handbook. Minneapolis: University of Minnesota Press.

DeFries, J.C., & Fulker, D.W.(1988). Multiple regression analysis of twin data. Behavior Genetics, 16, 467-473.

DeFries, J.C., & Fulker, D.W.(1985). Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. Acta Geneticae Gemellologicae, 37, 205-216.

Deutsch, C.K., & Kinsbourne, M.(1990). Genetics and biochemistry in attention deficit disorder. In M. Lewis and S. Miller(eds). Handbook of Developmental Psychopathology, (pp.65-76). New York: Plenum.

Dong, H.K.(1985). Non-gramian and singular matrices in maximum likelihood factor analysis. Applied Psychological Measurement, 9, 363-366.

Dunn, J.(1983). Sibling relationships in early childhood. Child Development, 54, 787-811.

Dunn, J., & Kendrick, C.(1981). Social behavior of young siblings in the family context: differences between same-sex and different-sex dyads. Child Development, 52, 1265-1273.

Dunn, J., & McGuire, S.(1992). Sibling and peer relationships in childhood. Journal of Child Psychology and Psychiatry, 33, 67-105.

Earls, F.(1980). The prevalence of behavior problems in 3-year-old children: Comparison of the reports of mothers and fathers. Journal of the American Academy of Child Psychiatry, 19, 439-452.

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

Eaves, L.J. (1986). Dominance alone is not enough. Behavior Genetics, 16, 27-34.

Edelbrock, C., Rende, R., & Plomin, R. (1992). Genetic and environmental effects on competence and problem behaviors in childhood and early adolescence. paper presented at the anual meeting of the Behavior Genetic Association.

Eysenck, H.J.(1967). The Biological Basis of Personality. Springfield: Thomas.

Eysenck, S.B.G.(1969). Personality dimensions in children . In H.J. Eysenck, and S.B.G. Eysenck. Personality Structure and Measurement(pp 265-317). London: Paul.

Falconer, D.S.(1989). Introduction to Quantitative Genetics. Essex: Longman Scientific and Technical.

Field, T., & Greenberg, R.(1982). Temperament ratings by parents and teachers of infants, toddlers, and preschool children. Child Development, 53, 160-163.

Fisher, R.A.(1918). The correlation between relatives on the supposition of Menelian inheritance. Translations of the Royal Society, Edingburgh, 52, 399-433.

Folstein, S., & Rutter, M.(1977). Infantile autisme: a genetic study of twin pairs of 21 twin pairs. Journal of Child Psychology and Psychiatry, 18, 297-321.

Fowler, P. O., & Park, R. M. (1979). Factor structure of the preschool behavior questionnaire in a normal population. Psychological Reports, 45, 599-606.

Gau, J.S., Silberg, J.L., Erickson, M.T., & Hewitt, J.K.(1992). Childhood behavior problems: A comparison of twin and non-twin samples. Acta Geneticae Medicae et Gemellologiae, 41, 53-63.

Gerbing, D.W., & Anderson, J.C.(1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. Psychometrica, 52, 99-111.

Ghodsian-Carpey, J., & Baker, L.A.(1987). Genetic and Environmental influences on aggression in 4- to 7-year-old twins. Aggressive Behavior, 13, 173-186.

Goldsmith, H.H.(1983). Genetic influences on Personality from infancy to adulthood. Child Development, 54, 331-355.
Goldsmith, H.H.(1989). Behavior-Genetic Approaches to Temperament. In Kohnstamm, G.A., Bates, J.E. and Rothbart. M.K.(eds), Temperament in Childhood. New York: Wiley.
Goldsmith, H.H., & Gottesman, I.I.(1981). Origins of variation in behavioral style: A longitudinal study of temperament in young twins. Child Development, 52, 91-103.
Goldstein, H.(1987). Multilevel Models in Educational and Social Research. New York: Oxford University Press.
Goodman, R., & Stevenson, J.(1989a). A twin study of hyperactivity-I. An examination of hyperactivity scores and categories derived from Rutter teacher and parent questionnaires. Journal of Child Psychology and Psychiatry, 30, 671-689.
Goodman, R., & Stevenson, J.(1989b). A twin study of hyperactivity-II. The aetiological role of genes, family relationships and perinatal adversity. Journal of Child Psychology and Psychiatry, 30, 691-709.
Gorsuch, R. L.(1983). Factor Analysis. Hillsdale, NJ: Erlbaum.
Gottesman, I.I.(1963). Heritability of personality: A demonstration. Psychological Monographs, 77, (No. 9, whole No. 572).
Gottesman, I.I.(1965). Personality and natural selection. In S.G. Vandenberg(ed), Methods and Goals in Human Behavioral Genetics, (pp. 63-81). New York: Academic Press.
Graham, P., & Rutter, M.(1985). Adolescent disorders. In M. Rutter and L. Hersov(eds), Child and Adolescent Psychiatry: Modern Approaches, (351-368). Oxford: Blackwell Scientific.
Graham, P., & Stevenson, J.(1985). A twin study of genetic influences on behavioral deviance. Journal of the American Academy of Child Psychiatry, 24, 33-41.
Hagnell, O., & Kreitman, N.(1974). Mental illness in married pairs in a total population. British Journal of Psychiatry, 125, 293-302.
Hay, D.A., & O'Brien, P.J.(1984). The role of parental attitudes in the development of temperament in twins at home, school, and in test situations. Acta Geneticae Medicae et Gemellologiae, 33, 191-204.
Hay, D.A., & O'Brien, P.J.(1987). Early influences on the school adjustment of twins. Acta Geneticae Medicae et Gemellologiae, 36, 213-222.
Hays, W.L.(1981). Statistics. Tokyo: Holt-Saunders.
Heath, A.C., Neale, M.C., Hewitt, J.K., Eaves, L.J., & Fulker, D.W.(1989). Testing structural equation models for twin data using LISREL. Behavioral Genetics, 19, 9-35.
Hersov, L.(1985). Emotional disorders. In M. Rutter and L. Hersov(eds), Child and Adolescent Psychiatry: Modern Approaches(2nd edn, pp.368-381). Oxford: Blackwell Scientific.
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, 3, 293-318.
Hoge, R. D., Meginbir, L., Khan, Y., & Weatherall, D.(1985). A multitrait-multimethod analysis of the preschool behavior questionnaire. Journal of Abnormal Child Psychology, 13, 119-127.
Holland, A.J., Hall, A., Murray R., Russel, G.F.M., & Crisp, A.H.(1984). Anorexia nervosa: a study of 34 twin pairs and one set of triplets. British Journal of Psychiatry, 145, 414-419.
James, A., & Taylor, E.(1990). Sex differences in the hyperkinetic syndrome of childhood. Journal of Child Psychology and Psychiatry, 31, 437-446.
Jary, M.L., & Stewart, M.A.(1985). Psychiatric disorder in the parents of adopted children with aggressive conduct disorder. Neuropsychobiology, 13, 7-11.
Jöreskog, K.G.(1971). Statistical analysis of sets of congeneric tests. Psychometrica. 36 ,109-133.
Jöreskog, K.G.(1978). Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.
Jöreskog, K.G., & Sörbom, D.(1988). PRELIS: A Preprocessor for LISREL. Mooresville, Indiana: Scientific Software.

Jöreskog, K.G., & Sörbom, D.(1989). LISREL 7. A Guide to the Program and Applications. Chicago: SPSS Inc.

Judd, C.M., Smith, E.R., & Kidder, L.H.(1991). Research Methods In Social Relations. Orlando: Holt Rinehart and Winston.

Keller, M.B., & Frances, M.S.(1989). Dysthymic and cyclothymic disorders. In C.G. Last, and M. Hersen(eds), Handbook of Child Psychiatric Diagnosis,(330-343). New York: Wiley.

Kim, C.C., Dales, R.J., Connor, R., Walters, J. & Witherspoon, R.(1969). Social interactions of like sex twins and singletons in relation to intelligence, language and physical development. Journal of Genetic Psychology, 114, 203-214.

Kohn, M., & Rosman, B. L.(1972). A social competence and symptom checklist for the preschool child: Factor dimensions, their cross-instrument generality, and longitudinal persistence. Developmental Psychology, 6, 430-444.

Koot, H.M., Van Den Oord, E.J.C.G., Verhulst, F.C., Boomsma, D.I. Dimensions of problem behavior among young preschoolers: Factor structure of the Child Behavior Checklist/2-3. submitted.

Koot, H.M., & Verhulst, F.C.(1991). Prevalence of problem behavior in Dutch children aged 2-3. Acta Psychiatrica Scandinavica, 83, suplementum 367.

Lavori, P.W., Keller, M.B., Beardslee, W.R., & Dorer, D.J.(1988). Affective disorder in childhood: seperating the familial component of risk from individual characteristics of children. Journal of Affective Disorders, 15, 303-311.

Leckman, J.F., Weissman, M.M., Pauls, D.L., & Kidd, K.K.(1987). Family-genetic studies and identification of valid diagnostic categories in adult and child psychiatry. British Journal of Psychiatry, 151, 39-44.

Lee, C.L., & Bates, J.E.(1985). Mother-child interaction at age two years and perceived difficult temperament. Child Development, 56, 1314-1325.

Loehlin, J.C.(1989). Partitioning environmental and genetic contributions to behavioral development. American Psychologist, 44, 1285-1292.

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

Livingston, R., Nugent, H., Rader, L., & Smith, G.R.(1985). Family histories of depressed and severely anxious children. American Journal of Psychiatry, 142, 1497-1499.

Lytton, H.(1980). Parent-Child Interaction: The Socialization Process Observed in Twin Singleton Families. New York: Plenum.

Lytton, H., Conway, D., & Sauvé(1977). The impact of twinship on parent-child interaction. Journal of Personality and Social Psychology, 35, 97-107.

MacCallum, R.(1986). Specification searches in covariance structure modeling. Psychological Bulletin, 9, 225-244.

Marsh, H.W., Balla, J.R., & McDonald, R.P.(1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin. 103, 391-410.

Martin, B., & Hoffman, J.A.(1990). Conduct disorders. In M. Lewis and S. Miller(eds). Handbook of Developmental Psychopathology, (pp.109-119). New York: Plenum.

Martin, N.G., & Eaves, L.J.(1977). The genetical analysis of covariance structure. Heredity, 38, 79-95.

Martin, N.G., Eaves, L.J., Kearsey, M.J., & Davies, P.(1978). The power of the classical twin study. Heredity, 40, 97-116.

Mascie-Taylor, C.G.N.(1989). Spouse similarity for IQ and personality and convergence. Behavior genetics, 19, 223-227.

Matheny, A.P., & Dolan, A.B.(1980). A twin study of Personality and temperament during middle childhood. Journal of Research in Personality, 14, 224-234.

Matheny, A.P., Wilson, R.S., Dolan, A. Brown., & Krantz, J.Z.(1981). Behavioral contrasts in twinship: Stability and patterns of differences in childhood. Child Development, 52, 579-

588.
Mather, K., & Jinks, J.L.(1971). Biometrical Genetics. London: Chapman and Hall.
McArdle, J.J., & Goldsmith, H.H.(1990). Alternative common factor models for multivariate biometric analyses. Behavior Genetics, 20, 569-609.
McGuire, J., & Richman, N.(1986). Screening for behaviour problems in nurseries: The reliability and validity of the Preschool Behaviour Checklist. Journal of Child Psychology and Psychiatry, 24, 273-280.
McGuffin, P.(1987). The new genetics and childhood psychiatric disorder. Journal of Child Psychology and Psychiatry, 28, 215-222.
McGuffin, P., Gottesman, I.I.(1985). Genetic influences on normal and abnormal development. In M. Rutter and L. Hersov(eds), Child and Adolescent Psychiatry: Modern Approaches, (2nd edn, pp.17-33). Oxford: Blackwell Scientific.
Merikangas, K.R., Weissman, M.M., Prusoff, B.A., & John, K.(1988). Assortive mating and affective disorders: psychopathology in offspring. Psychiatry, 51, 48-58.
Morrison, J.R., & Stewart, M.A.(1971). A family study of the hyperactive child syndrome. Biological Psychiatry, 3, 189-195.
Morrison, J.R., & Stewart, M.A.(1973). The psychiatric status of the legal families of adopted hyperactive children. Archives of General Psychiatry, 28, 888-891.
Muthén, B., & Kaplan, D.(1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. British Journal of Mathematical and Statistical Psychology, 38, 171-189.
Neale, M.C., & Cardon, L.R.(1992). Methodology for Genetic Studies of Twins and Families. Dordrecht: Kluwer.
Neale, M.C., & Martin, N.G.(1989). The effects of age, sex, and genotype on self-report drunkennes following a challenge dose of alcohol. Behavioral Genetics, 19, 63-78.
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.
Newman, H.H., Freeman, F.N., & Holzinger, K.J.(1966). Twins: A Study of Heredity and Environment. Chicago: University of Chicago Press.
Nichols, P.L., & Chen, T.(1981). Minimal Brain Dysfunction: A Prospective Study. Hillsdale, NJ: Lawrence Erlbaum.
O'Connor, M., Foch, T., Sherry, T., & Plomin, R.(1980). A twin study of specific behavioral problems of socialization as viewed by parents. Journal of Abnormal Child Psychology, 8, 189-199.
Owen, D., & Sines, J.O.(1970). Heritabil;ity of personality in children. Behavior Genetics, 1, 235-248.
Pardes, H., Kaufmann, C.A., Pincus, H.A., & West, A.(1989). Genetics and psychiatry: Past discoveries, current dilemmas, and future directions. The American Journal of Psychiatry, 146, 435-443.
Parke, R. D., & Slaby, R. G. (1983). The development of aggression. In E. M. Hetherington (Ed), Socialization, personality, and social development (Vol. 4, pp. 547-642). In P. Mussen (Series Ed), Handbook of Child Psychology. New York: Wiley.
Patterson, G.R.(1982). Coercive Family Process. Eugene, Oregon: Castalia Publishing Company.
Pauls, D.L., Cohen, D.J., Heimbuch, R., Detlor, J., & Kidd, K.K. (1981). Familial patterns and transmission of Gilles de la Tourette Syndrome and multiple tics. Archives of general psychiatry, 38, 1091-1093.
Phillips, K., Fulker, D.W., Carey, G., & Nagoshi (1987). Direct marital assortment for cognitive and personality variables. Behavior genetics, 17, 347-356.
Plomin, R.(1986). Development, Genetics and Psychology. Hillsdale, NJ: Lawrence Erlbaum Associates.

Plomin, R.(in press). Genetic risk and psychosocial disorders. In M. Rutter and P. Casaer(eds), Biological Risk Factors for Psychosocial Disorder. Cambridge: Cambridge University.
Plomin, R., & Daniels, D.(1986). Genetics and shyness. in W.H. Jones, J.M. Cheek and S.R. Briggs(Eds), Shyness: Perspectives on Research and treatment, 63-80.
Plomin, R. and Daniels, D.(1987). Why are children in the same family so different from one another?. Behavioral and Brain Sciences, 10, 1-60.
Plomin, R., DeFries, J.C., & Fulker, D.W.(1988). Nature and Nurture during infancy and early childhood. Cambridge University press.
Plomin, R., DeFries, J.C., & McClearn, G.E.(1990). Behavioral Genetics: A Primer. San Fran sisco: Freeman.
Plomin, R., & Foch, T.T.(1980). A twin study of objectively assessed personality in childhood. Journal of Personality and Social Psychology, 39, 680-688.
Plomin, R., Foch, T.T., & Rowe, D.C.(1981). Bobo clown agression in childhood: Environment not genes. Journal of Research in Personality, 15, 331-342.
Plomin, R., Nitz, K., & Rowe, D.C.(1990). Behavior genetics and aggressive behavior in childhood. In M.Lewis and S. Miller(eds). Handbook of developmental psychopathology(pp.119-133). New York: Plenum.
Plomin, R., & Rende, J.L.(1991). Human behavioral genetics. Annual Review of Psychology, volume 42. In press.
Plomin, R., Rende, R.D., & Rutter, M.L.(in press). Quantitative genetics and developmental psychopathology. In D. Cicchetti and S. Toth(eds.), Rochester symposium on developmental psychopathology, vol. 2: Internalizing and Externalizing Expressions of Dysfunction. Hillsdale, NJ: Erlbaum.
Poznanski, E.O., Cook, S.C., & Carroll, B.J.(1979). A depression rating scale for children. Pediatrics, 64, 442-450.
Price, R.A., & Vandenburg, S.G.(1980). Spouse similarity in American and Swedish couples. Behavior genetics, 10, 59-71.
Puig-Antich, J., Goetz, D., Davies, M., Kaplan, T., Davies, S., Ostrow, L., Asnis, L., Twomey, J., Iyengar, S., & Ryan, N.D.(1989). A controlled family history study of prepubertal major depressive disorder. Archives of General Psychiatry, 46, 406-419.
Quay, H. C.(1986). Classification. In H. C. Quay, & J. S. Werry(Eds), Psychopathological Disorders of Childhood (pp.1-34). New York: Wiley.
Quay, H.C., and Werry, J.S.(1986). Psychopathological Disorders of Childhood. New York: Wiley.
Quinton, D., & Rutter, M.(1984). Family pathology and child psychiatric disorder: A four-year prospective study. In A.R. Nicol(ed), Longitudinal Studies in Child Psychology and Psychiatry: Practical Lessons from Research Experience(91-134). New York: Wiley.
Rachman, S.(1969). Extraversion and neuroticisme in childhood. In H.J. Eysenck, and S.B.G. Eysenck. Personality Structure and Measurement(pp 253-265). London: Paul.
Richman, N. & Lansdown, R. (Eds). Problems of preschool children. Chichester: Wiley, 1988.
Richman, N., Stevenson, J. & Graham, P.J.(1982). Preschool to School: A Behavioral Study. New York: Academic Press.
Rosenbaum, J.F., Biederman, J., Gersten, M., Hirshfeld, D.R., Meminger, S.R., Herman, J.B., Kagan, J., Reznick, S., & Snidman, N.(1988). Behavioral inhibition in children of parents with panic disorder and agoraphobia. Archives of General Psychiatry, 45, 463-471.
Rowe, D.C.(1983). Biometrical genetic models of self-reported delinquent behavior: A twin study. Behavior Genetics, 13, 473-490.
Rutter, M.(1977). Influences on development. In M. Rutter and L. Hersov(eds), Child and Adolescent Psychiatry: Modern Approaches. Oxford: Blackwell Scientific.
Rutter, M.(1988). Epidemiological approaches to developmental psychopathology. Archives of General Psychiatry, 486-495.

Rutter, M.(1991). Nature, nurture and psychopathology: A new look at an old topic. Develop ment and Psychopathology, 3, 125-136.
Rutter, M., Bolton, P., Harrington, R., Le Couteur, A., Macdonald, H., & Simonoff, E. (1990a). Genetic factors in child psychiatric disorders-I. A review of research strategies. Journal of child psychology and psychiatry, 31, 3-37.
Rutter, M., & Cox, A.(1985). Other family influences. In M. Rutter and L. Hersov(eds), Child and Adolescent Psychiatry: Modern Approaches(2nd edn, pp. 58-81). Oxford: Blackwell Scientific.
Rutter, M., Graham, P., Chadwick, O., and Yule, W.(1976). Adolescent turmoil: fact or fiction? Journal of Child Psychology and Psychiatry, 17, 35-56.
Rutter, M., & Hemming, M.(1981). In M. Rutter, J. Tizard, and K. Whitmore(eds). Education, Health, and Behaviour. Melbourne: Krieger.
Rutter, M., Macdonald, H., Le Couteur, A., Harrington, R., Bolton, P. and Bailey, A.(1990b). Genetic factors in child psychiatric disorders-II. Empirical findings. Journal of Child Psychology and Psychiatry, 31, 39-83.
Rutter, M., & Redshaw, J.(1991). Annotation: Growing up as a twin: Twin-singleton differences in psychological development. Journal of Child Psychology and Psychiatry, 32, 885-895.
Rutter, M., Tizard, J., & Whitmore, K.(1970). Education, Health, and behaviour. London: Longman.(reprinted 1981, Melbourne Krieger).
Safer, D.J.(1973). A familial factor in minimal brain dysfunction. Behavior Genetics, 3, 175-187.
SAS Institute(1989). SAS User's Guide, Release 6.07.02. Cary, NC: SAS Institute.
Scarr, S.(1966). Genetic factors in activity motivation. Child Development, 37, 663-673.
Shaffer, D.(1977). Depression, mania and suicidal acts. In M. Rutter and L. Hersov(eds), Child and Adolescent Psychiatry: Modern Approaches, (698-720). Oxford: Blackwell Scientific.
Shields, J.(1977). Polygenic influences. In M. Rutter and L. Hersov(eds), Child and Adolescent Psychiatry: Modern Approaches, (pp.22-46). Oxford: Blackwell Scientific.
Silvia, E.S.M., & MacCallum, R.C.(1988). Some factors affecting the success of specification searches in covariance structure modeling. Multivariate Behavioral Research, 23, 297-326.
SPSS(1986). SPSS/PC+: For the IBM/XT/AT. Chigago: SPSS inc.
Stevenson, J., Batten, N., & Cherner, M.(1992). Fears and fearfulness in children and adoles cents: A genetic analysis of twin data. Journal of Child Psychology and Psychiatry, 33, 977-986.
Tanaka, J. S., & Huba, G. J. (1984). Hierarchical confirmatory factor analyses of psychological distress measures. Journal of Personality and Social Psychology, 46, 621-635.
Taylor, E.(1985). Syndromes of overactivity and attention deficit. In M. Rutter, and L. Hersov, L.(eds). Child and Adolescent Psychiatry: Modern Approaches, (424-444). Oxford: Blackwell Scientific.
Taylor, E.(1988). Attention deficit and conduct disorder syndromes. In M. Rutter, A. H. Tuma, & I. S. Lann(Eds.), Assessment and Diagnosis in Child Psychopathology (pp. 377-407). New York: Guilford Press.
Thomas, A., Chess, S., & Birch, H.G.(1968). Temperament and Behavior Disorders in Children. New York: New York University Press.
Thompson, W.R.(1966). Multivariate experiment in behavior genetics. In R.B. Cattell(ed) Handbook of Multivariate Experimental Psychology(711-731). Chigago: Rand McNally.
Tizard, B.(1977). Adoption: a Second Chance. London: Open Books.
Torgersen, A.M. (1982). Influence of genetic factors on temperament development in early childhood. In Temperamental Differences in Infants and Young Children(Ciba Foundation symposium 89, pp. 141-154). London: Pitman.
Trad, P. V.(1989). The Preschool Child. New York: Wiley.
Tremblay, R. E., Desmarais-Gervais, L., Gagnon, C., & Charlebois, P.(1987). The Preschool Behaviour Checklist: Stability of its factor structure between cultures, sexes, ages and so-

cioeconomic classes. International Journal of Behavioral Development, 10, 467-484.
Tucker, L.R., & Lewis C.(1973). The reliability coefficient for maximum likelihood factor analysis. Psychometrica, 38, 1-10.
Vandenberg, S.G.(1962). The hereditary abilities study: Hereditary components in a psychological test battery. American Journal of Human Genetics, 14, 220-237.
Vandenberg, S.G., & Crowe, L. (1989). Genetic factors in childhood psychopathology, In B.B. Lahey and A.E. Kazdin(eds), Advances in Clinical Child Psychology(vol. 12, pp.139-176). New York: Plenum.
Vandenberg, S.G., Singer, S.M., & Pauls, D. (1986). The Heredity of behavior Disorders in Adults and Children. New York: Plenum.
Van Den Oord, E.J.C.G., Boomsma, D.I., & Verhulst, F.C.(1992). A study of problem behaviors in 10- to 15-year-old biologically related and unrelated international adoptees. submitted.
Van Den Oord, E. J. C. G., Verhulst, F. C., Boomsma, D. I., & Orlebeke, J. F. (1992). A multivariate genetic analysis of problem behaviors in three-year-old twins. submitted.
Van Den Oord, E.J.C.G., Verhulst, F.C., & Boomsma.(1992). Psychometric properties of Achenbach's cross-informant syndrome constructs in a sample of international adoptees. submitted.
Van Den Oord, E. J. C. G., Koot, H. M., Boomsma, D. I., Verhulst, F. C., & Orlebeke, J. F.(1993). A twin-singleton comparison of problem behavior in 2-3-year olds. submitted.
Van Den Oord, E. J. C. G., Verhulst, F. C., Boomsma, D. I., & Orlebeke, J. F. (1992). A genetic study of maternal and paternal ratings of problem behaviors in three-year-old twins. submitted.
Van Driel, O.P. (1978). On various causes of improper solutions in maximum likelihood factor analysis. Psychometrica, 43, 225-243.
Van Westerlaak, J.M., Kropman, J.A., & Collaris, J.W.M.(1975). Beroepenklapper [Index of occupations]. Nijmegen: Instituut voor Toegepaste Sociologie.
Verhulst, F.C., Akkerhuis, G.W., & Althaus, M. (1985). Mental health in dutch children: I a cross-cultural comparison. Acta Psychiatrica Scandinavica, 72, Suppl. 323.
Verhulst, F.C, Althaus, M., & Versluis-den Bieman, H.J.M.(1990a). Problem behavior in inter national adoptees: I. An epidemiological study. Journal of the American Academy of Child and Adolescent Psychiatry, 29, 94-103.
Verhulst, F.C., Althaus, M., & Versluis-den Bieman, H.J.M.(1990b). Problem behavior in inter national adoptees: II. Age at placement. Journal of the American Academy of Child and Adolescent Psychiatry, 29, 104-111.
Verhulst, F.C., Althaus, M., & Versluis-den Bieman, H.J.M.(1991). Damaging backgrounds: Later adjustment of international adoptees. Journal of the American Academy of Child and Adolescent Psychiatry, 31, 518-525.
Verhulst, F.C., & Koot, H.M. (1992). Child Psychiatric Epidemiology: Concepts, Methods, and Findings. Developmental Clinical Psychology and Psychiatry, vol. 23. Newbury Park: Sage.
Verhulst, F.C., & Van Der Ende, J.(1991). Assessment of child psychopathology: relationships between different methods, different informants and clinical judgement of severity. Acta Psychiatrica Scandinavica, 84, 155-159.
Verhulst, F.C., & Van Der Ende, J.(1992a). Six-Year Developmental Course of Internalizing and Externalizing Problem Behaviors. Journal of the American Academy of Child and Adolescent Psychiatry, 31, 924-931.
Verhulst, F.C., & Van Der Ende, J.(1992b). Six-year stability of Parent reported Problem behavior in an epidemiological sample. Journal of Abnormal Child Psychology, 20, 595-610.
Verhulst, F.C., & Van Der Ende.(1993). 'Comorbidity' in an epidemiological: a longitudinal

perspective. Journal of Child Psychology and Psychiatry. in press.
Verhulst, F.C., Versluis-den Bieman, H.J.M., Van Der Ende, J., Berden, G.F.M.G., & Sanders-Woudstra, J.A.R.(1990). Problem behavior in international adoptees: III. Diagnosis and child psychiatric disorders. Journal of the American Academy of Child and Adolescent Psychiatry, 29, 420-428.
Weissman, M.M., Merikangas, K.R., Wickramaratne, P., Kidd, K.K., Prusoff, B.A., Leckman, J.F., & Pauls, D.L.(1986). Understanding the clinical heterogeneity of major depression using family data. Archives of General Psychiatry, 43, 430-434.
Weissman, M.M., Warner, V., Wickramaratne, P., & Prusoff, B.A.(1988). Early-onset major depression in parents and their children. Journal of Affective Disorders, 15, 269-277.
Weissman, M.M., Wickramaratne, P., Merikangas, K.R., Leckman, J.F., Prusoff, B.A., Caruso, K.A., Kidd, K.K., & Gammon, D.(1984). Onset of major depression in early adulthood. Archives of General Psychiatry, 41, 1136-1143.
Welner, Z., Welner, A., Stewart, M., Palkes, M.A., & Wish, E. (1977). A controlled study of siblings of hyperactive children. Journal of Nervous and Mental Disorders, 165, 110-117.
Weng, J.L.J.(1990). Aspects of Covariance Structure Analysis with Dependent Observations. Doctoral dissertation. University of California, Los Angeles.
Geneva: Author.
Wierzbicki, M.(1987). Similarity of monozygotic and dizygotic child twins in level and lability of subclinically depressed mood. American Journal of Orthopsychiatry, 57, 33-40.
Willerman, L.(1973). Activity level and hyperactivity in twins. Child Development, 44, 288-293.
Wilson, R.S., Brown, A.M., & Matheny, A.P.(1971). Emergence and persistence of behavioral differences in twins. Child Development, 42, 1381-1398.
World Health Organization (1989). ICD-10, 1989 Draft of chapter V, Categories F00-F99: Mental and behavioral disorders (including disorders of psychological development).
Young, P.A., Eaves, L.J., & Eysenck, H.J.(1980). Intergenerational stability and change in the causes of variation in adult and juvenile personality. Journal of Personality and individual differences, 1, 35-55.

Appendix 1[1]

CHILD BEHAVIOR CHECKLIST FOR AGES 4–18

For office use only
ID #

CHILD'S NAME

SEX ☐ Boy ☐ Girl

AGE

ETHNIC GROUP OR RACE

TODAY'S DATE Mo. _____ Date _____ Yr. _____

CHILD'S BIRTHDATE Mo. _____ Date _____ Yr. _____

GRADE IN SCHOOL _____

NOT ATTENDING SCHOOL ☐

Please fill out this form to reflect *your* view of the child's behavior even if other people might not agree. Feel free to write additional comments beside each item and in the spaces provided on page 2.

PARENTS' USUAL TYPE OF WORK, even if not working now. *(Please be specific—for example, auto mechanic, high school teacher, homemaker, laborer, lathe operator, shoe salesman, army sergeant.)*

FATHER'S TYPE OF WORK: _____

MOTHER'S TYPE OF WORK: _____

THIS FORM FILLED OUT BY:

☐ Mother (name): _____

☐ Father (name): _____

☐ Other—name & relationship to child: _____

I. Please list the sports your child most likes to take part in. For example: swimming, baseball, skating, skate boarding, bike riding, fishing, etc.

☐ None

	Compared to others of the same age, about how much time does he/she spend in each?				Compared to others of the same age, how well does he/she do each one?			
	Don't Know	Less Than Average	Average	More Than Average	Don't Know	Below Average	Average	Above Average
a. _____	☐	☐	☐	☐	☐	☐	☐	☐
b. _____	☐	☐	☐	☐	☐	☐	☐	☐
c. _____	☐	☐	☐	☐	☐	☐	☐	☐

II. Please list your child's favorite hobbies, activities, and games, other than sports. For example: stamps, dolls, books, piano, crafts, cars, singing, etc. (Do not include listening to radio or TV.)

☐ None

	Compared to others of the same age, about how much time does he/she spend in each?				Compared to others of the same age, how well does he/she do each one?			
	Don't Know	Less Than Average	Average	More Than Average	Don't Know	Below Average	Average	Above Average
a. _____	☐	☐	☐	☐	☐	☐	☐	☐
b. _____	☐	☐	☐	☐	☐	☐	☐	☐
c. _____	☐	☐	☐	☐	☐	☐	☐	☐

III. Please list any organizations, clubs, teams, or groups your child belongs to.

☐ None

	Compared to others of the same age, how active is he/she in each?			
	Don't Know	Less Active	Average	More Active
a. _____	☐	☐	☐	☐
b. _____	☐	☐	☐	☐
c. _____	☐	☐	☐	☐

IV. Please list any jobs or chores your child has. For example: paper route, babysitting, making bed, working in store, etc. (Include both paid and unpaid jobs and chores.)

☐ None

	Compared to others of the same age, how well does he/she carry them out?			
	Don't Know	Below Average	Average	Above Average
a. _____	☐	☐	☐	☐
b. _____	☐	☐	☐	☐
c. _____	☐	☐	☐	☐

1-91 Edition

V. 1. About how many close friends does your child have? (Do not include brothers & sisters) ☐ None ☐ 1 ☐ 2 or 3 ☐ 4 or more

2. About how many times a week does your child do things with any friends outside of regular school hours? (Do not include brothers & sisters) ☐ Less than 1 ☐ 1 or 2 ☐ 3 or more

VI. Compared to others of his/her age, how well does your child:

	Worse	About Average	Better	
a. Get along with his/her brothers & sisters?	☐	☐	☐	☐ Has no brothers or sisters
b. Get along with other kids?	☐	☐	☐	
c. Behave with his/her parents?	☐	☐	☐	
d. Play and work by himself/herself?	☐	☐	☐	

VII. 1. For ages 6 and older—performance in academic subjects. If child is not being taught, please give reason _____

	Failing	Below average	Average	Above average
a. Reading, English, or Language Arts	☐	☐	☐	☐
b. History or Social Studies	☐	☐	☐	☐
c. Arithmetic or Math	☐	☐	☐	☐
d. Science	☐	☐	☐	☐
Other academic subjects—for example: computer courses, foreign language, business. Do *not* include gym, shop, driver's ed., etc. e. _____	☐	☐	☐	☐
f. _____	☐	☐	☐	☐
g. _____	☐	☐	☐	☐

2. Is your child in a special class or special school? ☐ No ☐ Yes—what kind of class or school?

3. Has your child repeated a grade? ☐ No ☐ Yes—grade and reason

4. Has your child had any academic or other problems in school? ☐ No ☐ Yes—please describe

When did these problems start?

Have these problems ended? ☐ No ☐ Yes—when?

Does your child have any illness, physical disability, or mental handicap? ☐ No ☐ Yes—please describe

What concerns you most about your child?

Please describe the best things about your child:

Below is a list of items that describe children and youth. For each item that describes your child **now or within the past 6 months**, please circle the **2** if the item is **very true** or **often true** of your child. Circle the **1** if the item is **somewhat** or **sometimes true** of your child. If the item is **not true** of your child, circle the **0**. Please answer all items as well as you can, even if some do not seem to apply to your child.

0 = Not True (as far as you know) **1 = Somewhat or Sometimes True** **2 = Very True or Often True**

0 1 2 1. Acts too young for his/her age
0 1 2 2. Allergy (describe): _____
_____
0 1 2 3. Argues a lot
0 1 2 4. Asthma
0 1 2 5. Behaves like opposite sex
0 1 2 6. Bowel movements outside toilet
0 1 2 7. Bragging, boasting
0 1 2 8. Can't concentrate, can't pay attention for long
0 1 2 9. Can't get his/her mind off certain thoughts; obsessions (describe): _____
_____
0 1 2 10. Can't sit still, restless, or hyperactive
0 1 2 11. Clings to adults or too dependent
0 1 2 12. Complains of loneliness
0 1 2 13. Confused or seems to be in a fog
0 1 2 14. Cries a lot
0 1 2 15. Cruel to animals
0 1 2 16. Cruelty, bullying, or meanness to others
0 1 2 17. Day-dreams or gets lost in his/her thoughts
0 1 2 18. Deliberately harms self or attempts suicide
0 1 2 19. Demands a lot of attention
0 1 2 20. Destroys his/her own things
0 1 2 21. Destroys things belonging to his/her family or others
0 1 2 22. Disobedient at home
0 1 2 23. Disobedient at school
0 1 2 24. Doesn't eat well
0 1 2 25. Doesn't get along with other kids
0 1 2 26. Doesn't seem to feel guilty after misbehaving
0 1 2 27. Easily jealous
0 1 2 28. Eats or drinks things that are not food—*don't* include sweets (describe): _____
_____
0 1 2 29. Fears certain animals, situations, or places, other than school (describe): _____
_____
0 1 2 30. Fears going to school
0 1 2 31. Fears he/she might think or do something bad
0 1 2 32. Feels he/she has to be perfect
0 1 2 33. Feels or complains that no one loves him/her
0 1 2 34. Feels others are out to get him/her
0 1 2 35. Feels worthless or inferior
0 1 2 36. Gets hurt a lot, accident-prone
0 1 2 37. Gets in many fights
0 1 2 38. Gets teased a lot
0 1 2 39. Hangs around with others who get in trouble
0 1 2 40. Hears sounds or voices that aren't there (describe): _____
_____
0 1 2 41. Impulsive or acts without thinking
0 1 2 42. Would rather be alone than with others
0 1 2 43. Lying or cheating
0 1 2 44. Bites fingernails
0 1 2 45. Nervous, highstrung, or tense
0 1 2 46. Nervous movements or twitching (describe):
_____
_____
0 1 2 47. Nightmares
0 1 2 48. Not liked by other kids
0 1 2 49. Constipated, doesn't move bowels
0 1 2 50. Too fearful or anxious
0 1 2 51. Feels dizzy
0 1 2 52. Feels too guilty
0 1 2 53. Overeating
0 1 2 54. Overtired
0 1 2 55. Overweight
56. Physical problems without known medical cause:
0 1 2 a. Aches or pains (*not* headaches)
0 1 2 b. Headaches
0 1 2 c. Nausea, feels sick
0 1 2 d. Problems with eyes (describe): _____
_____
0 1 2 e. Rashes or other skin problems
0 1 2 f. Stomachaches or cramps
0 1 2 g. Vomiting, throwing up
0 1 2 h. Other (describe): _____
_____

Please see other side

0 = Not True (as far as you know) 1 = Somewhat or Sometimes True 2 = Very True or Often True

0 1 2 57. Physically attacks people
0 1 2 58. Picks nose, skin, or other parts of body (describe): _____
_____
0 1 2 59. Plays with own sex parts in public
0 1 2 60. Plays with own sex parts too much
0 1 2 61. Poor school work
0 1 2 62. Poorly coordinated or clumsy
0 1 2 63. Prefers being with older kids
0 1 2 64. Prefers being with younger kids
0 1 2 65. Refuses to talk
0 1 2 66. Repeats certain acts over and over; compulsions (describe): _____
_____
0 1 2 67. Runs away from home
0 1 2 68. Screams a lot
0 1 2 69. Secretive, keeps things to self
0 1 2 70. Sees things that aren't there (describe):
_____
_____
0 1 2 71. Self-conscious or easily embarrassed
0 1 2 72. Sets fires
0 1 2 73. Sexual problems (describe): _____
_____
_____
0 1 2 74. Showing off or clowning
0 1 2 75. Shy or timid
0 1 2 76. Sleeps less than most kids
0 1 2 77. Sleeps more than most kids during day and/or night (describe): _____
_____
0 1 2 78. Smears or plays with bowel movements
0 1 2 79. Speech problem (describe): _____
_____
0 1 2 80. Stares blankly
0 1 2 81. Steals at home
0 1 2 82. Steals outside the home
0 1 2 83. Stores up things he/she doesn't need (describe): _____
_____
0 1 2 84. Strange behavior (describe): _____
_____
0 1 2 85. Strange ideas (describe): _____
_____
0 1 2 86. Stubborn, sullen, or irritable
0 1 2 87. Sudden changes in mood or feelings
0 1 2 88. Sulks a lot
0 1 2 89. Suspicious
0 1 2 90. Swearing or obscene language
0 1 2 91. Talks about killing self
0 1 2 92. Talks or walks in sleep (describe): _____
_____
0 1 2 93. Talks too much
0 1 2 94. Teases a lot
0 1 2 95. Temper tantrums or hot temper
0 1 2 96. Thinks about sex too much
0 1 2 97. Threatens people
0 1 2 98. Thumb-sucking
0 1 2 99. Too concerned with neatness or cleanliness
0 1 2 100. Trouble sleeping (describe): _____
_____
0 1 2 101. Truancy, skips school
0 1 2 102. Underactive, slow moving, or lacks energy
0 1 2 103. Unhappy, sad, or depressed
0 1 2 104. Unusually loud
0 1 2 105. Uses alcohol or drugs for nonmedical purposes (describe): _____
_____
0 1 2 106. Vandalism
0 1 2 107. Wets self during the day
0 1 2 108. Wets the bed
0 1 2 109. Whining
0 1 2 110. Wishes to be of opposite sex
0 1 2 111. Withdrawn, doesn't get involved with others
0 1 2 112. Worries
113. Please write in any problems your child has that were not listed above:
0 1 2 _____
0 1 2 _____
0 1 2 _____

PLEASE BE SURE YOU HAVE ANSWERED ALL ITEMS.

UNDERLINE ANY YOU ARE CONCERNED ABOUT.

[1] reproduced by permission of T.M. Achenbach

Appendix 2[1]

CHILD BEHAVIOR CHECKLIST FOR AGES 2-3

For office use only
ID #

CHILD'S NAME

SEX ☐ Boy ☐ Girl

AGE

ETHNIC GROUP OR RACE

TODAY'S DATE

Mo.____ Day____ Yr.____

CHILD'S BIRTHDATE

Mo.____ Day____ Yr.____

Please fill out this form to reflect *your* view of the child's behavior even if other people might not agree about the behavior.

PARENT'S TYPE OF WORK *(Please be specific—for example, auto mechanic, high school teacher, homemaker, laborer, lathe operator, shoe salesman, army sergeant, even if parent does not live with child.)*

FATHER'S TYPE OF WORK:____

MOTHER'S TYPE OF WORK:____

THIS FORM FILLED OUT BY:

☐ Mother (name):____

☐ Father (name):____

☐ Other—name & relationship to child:____

Below is a list of items that describe children. For each item that describes the child **now or within the past 2 months**, please circle the 2 if the item is **very true or often true** of the child. Circle the 1 if the item is **somewhat or sometimes true** of the child. If the item is **not true** of the child, circle the 0. Please answer all items as well as you can, even if some do not seem to apply to the child.

0 = Not True (as far as you know) **1 = Somewhat or Sometimes True** **2 = Very True or Often True**

0	1	2	1.	Aches or pains (without medical cause)	0	1	2	33.	Feelings are easily hurt
0	1	2	2.	Acts too young for age	0	1	2	34.	Gets hurt a lot, accident-prone
0	1	2	3.	Afraid to try new things	0	1	2	35.	Gets in many fights
0	1	2	4.	Avoids looking others in the eye	0	1	2	36.	Gets into everything
0	1	2	5.	Can't concentrate, can't pay attention for long	0	1	2	37.	Gets too upset when separated from parents
0	1	2	6.	Can't sit still or restless	0	1	2	38.	Has trouble getting to sleep
0	1	2	7.	Can't stand having things out of place	0	1	2	39.	Headaches (without medical cause)
0	1	2	8.	Can't stand waiting; wants everything now	0	1	2	40.	Hits others
0	1	2	9.	Chews on things that aren't edible	0	1	2	41.	Holds his/her breath
0	1	2	10.	Clings to adults or too dependent	0	1	2	42.	Hurts animals or people without meaning to
0	1	2	11.	Constantly seeks help	0	1	2	43.	Looks unhappy without good reason
0	1	2	12.	Constipated, doesn't move bowels	0	1	2	44.	Angry moods
0	1	2	13.	Cries a lot	0	1	2	45.	Nausea, feels sick (without medical cause)
0	1	2	14.	Cruel to animals	0	1	2	46.	Nervous movements or twitching (describe):____
0	1	2	15.	Defiant					
0	1	2	16.	Demands must be met immediately					
0	1	2	17.	Destroys his/her own things	0	1	2	47.	Nervous, highstrung, or tense
0	1	2	18.	Destroys things belonging to his/her family or other children	0	1	2	48.	Nightmares
					0	1	2	49.	Overeating
0	1	2	19.	Diarrhea or loose bowels when not sick	0	1	2	50.	Overtired
0	1	2	20.	Disobedient	0	1	2	51.	Overweight
0	1	2	21.	Disturbed by any change in routine	0	1	2	52.	Painful bowel movements
0	1	2	22.	Doesn't want to sleep alone	0	1	2	53.	Physically attacks people
0	1	2	23.	Doesn't answer when people talk to him/her	0	1	2	54.	Picks nose, skin, or other parts of body (describe):____
0	1	2	24.	Doesn't eat well (describe):____					
0	1	2	25.	Doesn't get along with other children	0	1	2	55.	Plays with own sex parts too much
0	1	2	26.	Doesn't know how to have fun, acts like a little adult	0	1	2	56.	Poorly coordinated or clumsy
					0	1	2	57.	Problems with eyes without medical cause (describe):____
0	1	2	27.	Doesn't seem to feel guilty after misbehaving					
0	1	2	28.	Doesn't want to go out of home					
0	1	2	29.	Easily frustrated	0	1	2	58.	Punishment doesn't change his/her behavior
0	1	2	30.	Easily jealous	0	1	2	59.	Quickly shifts from one activity to another
0	1	2	31.	Eats or drinks things that are not food (describe):____	0	1	2	60.	Rashes or other skin problems (without medical cause)
					0	1	2	61.	Refuses to eat
0	1	2	32.	Fears certain animals, situations, or places (describe):____	0	1	2	62.	Refuses to play active games
					0	1	2	63.	Repeatedly rocks head or body
					0	1	2	64.	Resists going to bed at night

Please see other side

0 = Not True (as far as you know)	1 = Somewhat or Sometimes True	2 = Very True or Often True

0	1	2	65.	Resists toilet training (describe): _____	0	1	2	82.	Sudden changes in mood or feelings
					0	1	2	83.	Sulks a lot
0	1	2	66.	Screams a lot	0	1	2	84.	Talks or cries out in sleep
0	1	2	67.	Seems unresponsive to affection	0	1	2	85.	Temper tantrums or hot temper
0	1	2	68.	Self-conscious or easily embarrassed	0	1	2	86.	Too concerned with neatness or cleanliness
0	1	2	69.	Selfish or won't share	0	1	2	87.	Too fearful or anxious
0	1	2	70.	Shows little affection toward people	0	1	2	88.	Uncooperative
0	1	2	71.	Shows little interest in things around him/her	0	1	2	89.	Underactive, slow moving, or lacks energy
0	1	2	72.	Shows too little fear of getting hurt	0	1	2	90.	Unhappy, sad, or depressed
0	1	2	73.	Shy or timid	0	1	2	91.	Unusually loud
0	1	2	74.	Sleeps less than most children during day and/or night (describe): _____	0	1	2	92.	Upset by new people or situations (describe): _____
0	1	2	75.	Smears or plays with bowel movements	0	1	2	93.	Vomiting, throwing up (without medical cause)
0	1	2	76.	Speech problem (describe): _____	0	1	2	94.	Wakes up often at night
					0	1	2	95.	Wanders away from home
0	1	2	77.	Stares into space or seems preoccupied	0	1	2	96.	Wants a lot of attention
0	1	2	78.	Stomachaches or cramps (without medical cause)	0	1	2	97.	Whining
					0	1	2	98.	Withdrawn, doesn't get involved with others
0	1	2	79.	Stores up things he/she doesn't need (describe): _____	0	1	2	99.	Worrying
								100.	Please write in any problems your child has that were not listed above.
0	1	2	80.	Strange behavior (describe): _____	0	1	2		_____
					0	1	2		_____
0	1	2	81.	Stubborn, sullen, or irritable	0	1	2		_____

PLEASE BE SURE YOU HAVE ANSWERED ALL ITEMS.

UNDERLINE ANY YOU ARE CONCERNED ABOUT.

[1] reproduced by permission of T.M. Achenbach

Appendix 3a

Parameter estimates and fit statistics for models allowing 'general scalar sex limitation'.

	df.	χ^2	p.	h	c	e	s	scalar
Total score								
	15	11.11	.745	.532	.001	.555	.216	1.068
	15	11.61	.708	.475	.544/.405	.432	-	1.061
Internalizing grouping								
	15	12.50	.641	.000	.307	.549	.032	1.049
	15	12.50	.641	.000	.347/.336	.532	-	1.048
Withdrawn								
	15	17.50	.290	.000	.000	.655	.072	1.056
	15	17.00	.319	.000	.250/.096	.616	-	1.056
Somatic Complaints								
	15	28.26	.020	2.32	.001	6.83	.060	.927
	15	28.26	.020	2.11	2.54/2.76	6.44	-	.928
Anxious/depressed								
	15	10.22	.806	.114	.288	.610	.030	1.085
	15	10.20	.807	.134	.333/.313	.589	-	1.093
Social Problems								
	15	22.57	.094	3.52	.000	8.19	.093	1.139
	15	22.24	.102	3.60	3.87/2.55	7.31	-	1.138
Attention Problems								
	15	14.13	.516	.220	.043	.224	.018	1.104
	15	13.81	.540	.217	.079/.000	.217	-	1.104
Externalizing grouping								
	15	23.45	.075	.525	.156	.319	.057	1.226
	15	21.35	.126	.513	.278/.000	.285	-	1.228
Delinquent Behavior								
	15	25.33	.046	.531	.000	.623	.094	1.260
	15	24.47	.057	.527	.363/.042	.550	-	1.257
Aggressive Behavior								
	15	16.97	.321	.361	.125	.209	-.008	1.281
	15	15.85	.392	.353	.132/.000	.220	-	1.285

Note. Parameter h is additive genetic effect, c is shared environmental effect, e is non-shared environmental effect and s is sibling effect, - denotes parameter fixed at zero. First model for each scale allows sibling interaction, second model allows reduced shared environmental influences in singletons. Scalar girls is fixed at 1. Number before slash is appropriate to siblings, number after slash is appropriate to singletons.

Appendix 3b

Parameter estimates and fit statistics for models allowing 'specific scalar sex limitation'.

	df.	χ^2	p.	h_g	h_b	c_g	c_b	e_g	e_b	s	b
Total score											
	13	9.40	.742	.494	.552	.309	-.054	.506	.605	.209	*1.00*
	13	8.84	.785	.385	.592	.639	.487	.423	.443	-	.671
Internalizing grouping											
	13	12.45	.492	.000	.000	.343	.335	.531	.571	.013	*1.00*
	13	12.45	.491	.000	.000	.358	.351	.525	.564	-	.987
Withdrawn											
	13	16.52	.222	-.052	.202	.238	.001	.610	.663	.055	*1.00*
	13	16.30	.233	-.128	.258	.312	.201	.574	.617	-	.682
Somatic Complaints											
	13	18.44	.142	-.119	4.73	5.18	1.83	5.26	4.41	-.066	*1.00*
	13	17.11	.194	-.250	4.36	4.42	.524	5.61	5.03	-	1.37
Anxious/depressed											
	13	9.75	.715	.000	.000	.326	.200	.650	.665	.054	*1.00*
	13	9.62	.724	.153	.098	.288	.413	.606	.622	-	.851
Social Problems											
	13	21.37	.066	4.21	3.48	-.820	3.53	7.73	8.76	.087	*1.00*
	13	22.06	.054	4.77	3.51	2.96	5.23	7.00	8.09	-	.853
Attention Problems											
	13	13.97	.376	.228	.233	.068	.105	.215	.240	-.006	*1.00*
	13	13.81	.540	.221	.234	.066	.101	.219	.243	-	.000
Externalizing grouping											
	13	20.41	.085	.466	.718	.338	.035	.276	.297	.051	*1.00*
	13	16.19	.239	.444	.742	.420	.173	.263	.197	-	.000
Delinquent Behavior											
	13	24.42	.027	.445	.774	-.139	.208	.671	.652	.096	*1.00*
	13	23.91	.032	.393	.828	.410	.398	.626	.526	-	.000
Aggressive Behavior											
	13	13.94	.378	.309	.510	.264	.106	.174	.199	-.041	*1.00*
	16	10.52	.651	.319	.499	.256	.040	.183	.242	-	.000

Note. Parameter h is additive genetic effect, c is shared environmental effect, e is non-shared environmental effect, s is sibling effect, - denotes parameter fixed at zero. First model for each scale allows sibling interaction, second model allows reduced shared environmental influences in singletons. Parameter b is fixed at 1 for siblings and estimated for singletons. Subscript g refers to girls and subscript b refers to boys.

Appendix 4

Intercorrelations of CBCL/2-3 factors obtained from confirmatory factor analysis.

Clinical sample	1.	2.	3.	4.	5.	6.
1. Oppositional	1.000					
2. Withdrawn/Depr.	.431	1.000				
3. Aggressive	.681	.329	1.000			
4. Anxious	.401	.638	.158	1.000		
5. Overactive	.703	.390	.633	.242	1.000	
6. Sleep problems	.442	.121	.217	.320	.294	1.000
7. Somatic problems	.278	.350	.170	.354	.169	.307

Community sample	1.	2.	3.	4.	5.	6.
1. Oppositional	1.000					
2. Withdrawn/Depr.	.519	1.000				
3. Aggressive	.500	.332	1.000			
4. Anxious	.444	.534	.069	1.000		
5. Overactive	.712	.467	.465	.299	1.000	
6. Sleep problems	.285	.250	.162	.326	.293	1.000
7. Somatic problems	.344	.195	.079	.206	.254	.432

Twin sample	1.	2.	3.	4.	5.	6.
1. Oppositional	1.000					
2. Withdrawn/Depr.	.600	1.000				
3. Aggressive	.715	.487	1.000			
4. Anxious	.540	.638	.265	1.000		
5. Overactive	.754	.536	.684	.425	1.000	
6. Sleep problems	.489	.356	.372	.347	.395	1.000
7. Somatic problems	.351	.387	.232	.304	.277	.384

Appendix 5

Chi-squares for models testing differences in means and variances.

	baseline	girls=boys	MZ=DZ=Co.	MZ=DZ≠Co.	MZ≠DZ=Co.	DZ≠MZ=Co.
	df=6	df=9	df=10	if girls=boys df=10, if not df=8		
Total problem score						
mean	22.11(.00)	52.28(.00)	53.34(.00)	33.85(.00)	49.83(.00)	23.01(.00)
variance	5.57(.47)	7.28(.61)	18.42(.05)	7.87(.64)	16.09(.10)	17.08(.07)
Oppositional						
mean	6.68(.35)	9.85(.36)	27.84(.00)	21.47(.02)	23.84(.01)	9.85(.45)
variance	5.28(.26)	11.33(.25)	19.42(.04)	13.86(.18)	18.70(.05)	19.67(.03)
Depressed/Withdrawn						
mean	13.18(.04)	15.54(.08)	16.95(.08)	-	-	-
variance	6.84(.34)	10.77(.29)	10.09(.43)	-	-	-
Aggressive						
mean	11.22(.08)	192.70(.00)	29.73(.00)	27.55(.00)	15.62(.05)	16.83(.03)
variance	2.77(.838)	29.31(.00)	12.30(.27)	6.15(.63)	6.63(.58)	12.21(.14)
Anxious						
mean	9.72(.14)	16.63(.06)	22.67(.01)	21.88(.02)	17.07(.07)	22.47(.01)
variance	5.37(.50)	12.39(.19)	14.80(.14)	-	-	-
Overactive						
mean	3.58(.73)	26.33(.00)	28.13(.00)	10.82(.21)	24.33(.00)	9.28(.32)
variance	1.12(.98)	8.05(.53)	9.76(.46)	-	-	-
Sleep Problems						
mean	16.74(.01)	18.53(.03)	31.98(.00)	26.68(.00)	28.52(.00)	18.71(.05)
variance	4.56(.60)	7.63(.57)	11.63(.31)	-	-	-

Note. MZ refers to monozygote twins, DZ refers to dizygote twins, Co. refers to the community sample. Chi-squares that are underlined denote that the restrictions imposed by the model are appropriate (non-significant difference), probabilities are in parentheses, df denotes degrees of freedom.

For the total problem score, the Aggressive syndrome, and Sleep problems, the chi-squares indicated that the baseline model for the mean scores did not fit. The fit of the model could significantly be improved when the means of DZ same-sex twins were not constrained to be equal to the means for DZ opposite-sex twins. For these scales the means of DZ opposite-sex twins tended to be somewhat lower than for DZ same-sex twins.

Appendix 6

Observed twin variances (on diagonal), covariances (above diagonal), and correlations (below diagonal).

	MZ girls				MZ boys				DZ girls				DZ boys				Opposite sex			
	MRT1	FRT1	MRT2	FRT2	MRT1	FRT1	MRT2	MRT2	MRT1	FRT1	MRT2	FRT2	MRT1	FRT1	MRT2	FRT2	MRT1	FRT1	MRT2	FRT2
Total problem score																				
MRT1	.273	.183	.242	.185	.232	.192	.191	.148	.233	.138	.165	.099	.205	.161	.156	.132	.235	.161	.155	.085
FRT1	.784	.223	.175	.189	.799	.222	.150	.179	.632	.222	.094	.155	.739	.215	.128	.160	.718	.221	.117	.130
MRT2	.858	.688	.278	.205	.825	.674	.226	.169	.710	.439	.252	.158	.671	.509	.256	.210	.654	.523	.243	.170
FRT2	.759	.816	.809	.238	.656	.828	.761	.205	.449	.693	.707	.230	.530	.662	.785	.266	.376	.589	.736	.233
		(No.=127)				(No.=114)				(No.=142)				(No.=160)				(No.=241)		
Oppositional																				
MRT1	51.0	29.8	40.0	24.5	43.2	28.4	33.4	22.2	40.1	24.0	19.2	9.4	40.3	30.9	19.1	16.0	39.3	31.1	17.8	9.7
FRT1	.698	39.2	25.9	27.5	.694	38.5	21.0	26.2	.653	34.4	8.6	14.4	.725	41.6	13.4	19.0	.768	39.6	13.3	14.4
MRT2	.764	.578	52.2	32.7	.783	.555	41.6	21.0	.483	.241	39.6	23.6	.419	.274	52.6	40.2	.443	.324	42.2	29.8
FRT2	.605	.722	.762	36.4	.590	.732	.626	33.8	.258	.422	.664	34.1	.332	.403	.725	52.8	.250	.373	.742	37.7
		(No.=162)				(No.=142)				(No.=175)				(No.=200)				(No.=293)		
Depressed/Withdrawn																				
MRT1	.355	.151	.233	.149	.423	.273	.248	.191	.367	.125	.148	.059	.421	.216	.152	.094	.386	.147	.164	.080
FRT1	.469	.323	.102	.225	.655	.413	.177	.290	.367	.368	.082	.198	.525	.416	.057	.172	.450	.314	.092	.107
MRT2	.657	.318	.358	.180	.638	.462	.358	.176	.376	.215	.432	.257	.346	.131	.463	.261	.431	.272	.379	.183
FRT2	.425	.644	.496	.372	.476	.723	.474	.387	.154	.492	.611	.426	.222	.399	.594	.434	.247	.344	.548	.312
		(No.=164)				(No.=144)				(No.=174)				(No.=205)				(No.=290)		
Aggressive																				
MRT1	.079	.047	.062	.043	.091	.066	.068	.049	.062	.037	.035	.023	.085	.053	.043	.028	.062	.041	.033	.023
FRT1	.649	.069	.043	.054	.741	.082	.052	.060	.609	.061	.021	.027	.613	.083	.029	.038	.669	.061	.022	.029
MRT2	.800	.584	.075	.053	.776	.603	.087	.056	.537	.339	.071	.042	.522	.338	.082	.056	.454	.311	.085	.056
FRT2	.595	.776	.713	.069	.577	.755	.676	.077	.380	.455	.694	.058	.314	.451	.658	.084	.324	.403	.686	.083
		(No.=170)				(No.=146)				(No.=182)				(No.=207)				(No.=299)		
Anxious																				
MRT1	.075	.041	.056	.037	.057	.037	.032	.024	.067	.052	.016	.013	.069	.051	.023	.015	.059	.039	.014	.009
FRT1	.659	.061	.033	.045	.662	.053	.029	.033	.752	.071	.009	.022	.740	.069	.015	.020	.664	.062	.010	.015
MRT2	.714	.495	.081	.048	.560	.520	.058	.038	.239	.140	.066	.046	.317	.208	.073	.050	.234	.161	.059	.042
FRT2	.555	.689	.676	.070	.434	.624	.679	.054	.192	.319	.707	.068	.215	.285	.685	.071	.149	.249	.682	.060
		(No.=169)				(No.=146)				(No.=179)				(No.=207)				(No.=299)		
Overactive																				
MRT1	4.60	2.79	2.42	1.30	5.61	3.90	2.48	1.98	4.58	2.40	.48	-.29	5.03	3.18	.49	.45	4.63	3.33	.55	.16
FRT1	.681	3.99	1.73	1.86	.663	5.76	1.91	2.28	.562	4.30	-.27	.00	.666	4.63	.04	.52	.717	4.33	-.00	.14
MRT2	.531	.402	4.50	2.90	.468	.351	4.97	3.66	.102	-.059	4.74	3.42	.099	.008	4.82	3.89	.114	-.000	5.17	3.77
FRT2	.311	.456	.661	4.19	.351	.409	.694	5.37	-.061	.001	.692	5.03	.086	.101	.739	5.64	.031	.030	.719	5.03
		(No.=172)				(No.=151)				(No.=183)				(No.=206)				(No.=303)		
Sleep Problems																				
MRT1	.270	.203	.196	.144	.248	.138	.157	.086	.234	.142	.121	.091	.257	.190	.087	.075	.248	.179	.067	.046
FRT1	.800	.269	.157	.181	.590	.231	.102	.150	.651	.213	.082	.114	.718	.264	.080	.107	.714	.251	.054	.081
MRT2	.706	.570	.287	.197	.682	.479	.214	.144	.497	.363	.253	.172	.366	.327	.220	.166	.285	.237	.220	.135
FRT2	.579	.689	.724	.252	.372	.650	.673	.230	.374	.491	.681	.258	.288	.413	.684	.253	.211	.376	.684	.185
		(No.=161)				(No.=139)				(No.=178)				(No.=202)				(No.=299)		

Note. MRT1 is maternal rating of twin 1, FRT1 is paternal rating of twin 1, MRT2 is maternal rating of twin 2, FRT2 is paternal rating of twin 2. Twin 1 is the first born child. The average of the pairwise sample sizes was used as the number of observations (No.). Except for the Oppositional and Overactive syndromes, scores were log-transformed.

Appendix 7

Parameter estimates obtained from fitting the biometric, psychometric and bias model.

Biometric model

	h_m	h_f	h_{fm}	c_m	c_f	c_{fm}	e_m	e_f	e_{fm}
Total problem score									
girls	.295	.059	.240	.354	.268	.240	.183	.182	.083
boys	.302	.236	.237	.330	.179	.215	.193	.096	.158
Oppositional									
girls	4.592	1.101	4.637	3.268	2.031	.337	3.328	2.634	1.813
boys	5.614	1.026	5.250	2.078	1.455	-.888	2.995	2.976	1.182
Depressed/Withdrawn									
girls	.477	.223	.234	.182	.341	.115	.351	.333	.117
boys	.461	.245	.420	.222	.164	-.085	.383	.313	.107
Aggressive									
girls	.196	.138	.138	.118	.058	.072	.122	.111	.056
boys	.200	.148	.181	.155	.062	.071	.147	.122	.069
Anxious									
girls	.204	.054	.189	.064	.047	-.068	.151	.126	.074
boys	.191	.091	.176	.040	-.022	-.035	.163	.126	.082
Overactive									
girls	1.419	.901	.751	-.367	.231	-.361	1.575	1.211	1.166
boys	1.352	.694	.893	.452	.315	.606	1.757	1.385	1.259
Sleep									
girls	.282	.221	.211	.306	.157	.236	.287	.225	.236
boys	.391	.237	.263	.061	.132	.057	.272	.222	.171

Psychometric model

	h	c	e	h_m	h_f	c_m	c_f	e_m	e_f
Total score									
girls	.287	.257	.122	.082	.032	.234	.183	.136	.153
boys	.223	.315	.178	.146	.199	.174	.103	.183	.059
Oppositional									
girls	4.419	1.687	2.466	.806	.928	2.955	1.930	2.218	2.067
boys	5.294	-.108	1.897	.990	.919	2.716	1.686	2.218	2.067
Depressed/Withdrawn									
girls	.354	.066	.199	.295	-.102	.203	.309	.293	.283
boys	.414	.059	.208	.193	.253	.222	.188	.323	.255
Aggressive									
girls	.154	.110	.083	.105	.102	.074	.031	.089	.093
boys	.197	.089	.100	.066	.125	.113	-.021	.125	.099
Anxious									
girls	.183	-.023	.108	.083	.043	.070	.099	.104	.099
boys	.176	.033	.117	.005	.090	.075	.030	.116	.094
Overactive									
girls	1.059	-.347	1.344	.793	.605	.515	.059	.837	.987
boys	1.054	.567	1.501	.605	.496	.395	.177	.958	1.126
Sleep Problems									
girls	.233	.276	.219	.192	.152	.083	.135	.182	.178
boys	.316	.080	.217	.220	.155	.000	.128	.168	.176

Bias model

	h	c	e	b_m	b_f	u_m	u_f	a
Total score								
girls	.290	.256	.117	.250	.164	.142	.160	1.027
boys	.273	.326	.192	.136	.242	.040	.164	.812
Oppositional								
girls	4.275	1.641	2.372	3.204	1.794	2.404	1.942	1.071
boys	5.389	-.120	1.900	2.751	1.981	2.348	2.663	.968
Depressed/Withdrawn								
girls	.224	.147	.159	.388	.059	.377	.189	1.803
boys	.409	.095	.195	.267	.284	.352	.279	1.035
Aggressive								
girls	.175	.089	.075	.110	-.099	.105	.108	.968
boys	.184	.071	.083	.152	-.021	.133	.097	1.263
Anxious								
girls	.193	-.032	.113	.083	.119	.103	.109	.904
boys	.195	.038	.105	.098	.032	.134	.080	1.222
Overactive								
girls	1.045	-.347	1.287	.891	.367	1.010	.978	1.074
boys	1.037	.599	1.502	.610	.482	1.032	1.148	1.001
Sleep								
girls	.282	.291	.232	.094	.229	.169	.229	.817
boys	.390	.061	.257	-.009	.235	.088	.254	.687

Note. Parameter h is additive genetic effect, c is shared environmental effect, e is non-shared environmental effect, b is rater bias effect, and u is the square root of the unreliability. Parameter a is the loading from the paternal rating on the phenotype, the loading from the maternal rating is fixed at one. Subscript m refers to mother, subscript f refers to father.

Summary

The primary aim of this dissertation was to study genetic and environmental influences on problem behaviors in preschool children and adolescents. In chapter 1, it was argued that because of the continuous character of most child psychiatric conditions the methods of the quantitative genetic theory were appropriate, and that disorders should be assessed as quantitative variations of behavior rather than all-or-none categories. The dissertation was divided in two parts. The first part concerned genetic influences on problem behaviors in children and adolescents, and involved genetic analyses on a sample of 11- to 15-year-old international adoptees. The second part addressed problem behaviors in children of preschool age, and involved genetic analyses on a sample of 3-year-old twins.

Part 1. A short introduction to the methods that have been applied to study genetic influences on problem behaviors in children and adolescents, was presented in Chapter 2. This introduction was followed by a survey of findings from genetic studies of the commoner varieties of problem behaviors in children aged 4-18. The small sample sizes in most studies in this review, the different assessment procedures across studies, the use of possibly inappropriate genetic models, and the fact that estimates of genetic and environmental influences are population dependent, made it difficult to draw firm conclusions. To the extent that it was possible to draw general conclusions, it appeared that genetic influences were important to most problem behaviors. Evidence for shared environmental influences was found for antisocial behaviors. The too low DZ twin correlations found for social withdrawal and (hyper)activity, suggested that for these behaviors the commonly used model with additive genetic, shared environmental, and non-shared environmental influences may be inappropriate.

To obtain parental ratings of problem behaviors in the sample of international adoptees, the CBCL/4-18 (Child Behavior Checklist for Ages 4-18) was used. It was planned to use recently derived American CBCL/4-18 syndromes for the genetic analyses. Therefore, in chapter 3, the validity of these syndrome constructs was studied in the sample of international adoptees (N=2,148). Results were cross-validated on a clinical sample (N=1,387). Support was found for the validity of the constructs. However, in the adoption sample, the contribution of a number of items to the scales of the syndrome constructs was questionable. These items had very low variances, were not indicators of just one construct, or did not improve the reliability of the scale.

In chapter 4, American CBCL/4-18 syndromes, adapted to the Dutch sample of international adoptees, were used to study genetic and environmental influences on problem behaviors. The sample (mean age 12.43 years) comprised a group of biological siblings (111 pairs), a group of non-biological siblings (221 pairs), and a group of singletons (94). Non-shared environmental influences were most important. Genetic influences were substantial for externalizing behaviors, but unimportant for internalizing behaviors. For the CBCL total problem score, Attention Problems, and externalizing behaviors results were in agreement with findings from twin studies. The lack of genetic influences on internalizing behaviors was in contrast with results from twin studies. For the Externalizing grouping, Delinquent Behavior, and Aggressive Behavior, variances for singletons were significantly smaller than for siblings. Model fit indices indicated that these differences in variances are better attributed to smaller effects of factors associated with sibship size, than to active influences of siblings on each other. Significant sex differences were found for 7 of the 10 scales. The larger variances for boys on the Externalizing grouping and Aggressive Behavior were caused by genetic influences.

Part 2. For the genetic analyses on the sample of 3-year-old twins, the CBCL/2-3 (Child Behavior Checklist for Ages 2-3) was used to obtain parental ratings of problem behaviors. In chapter 5, the Dutch factor structure of the CBCL/2-3 investigated with three different samples - children referred to mental health services, children from the general population, and the sample of twin pairs. A series of exploratory and confirmatory factor analyses indicated a seven-factor model for all three samples. Syndromes were labeled Oppositional, Withdrawn/Depressed, Aggressive, Anxious, Overactive, Sleep Problems, and Somatic Problems. Internal consistency estimates, test-retest stability, and interparent agreement were moderate to high for the seven factors. Factor intercorrelations and a second-order factor analysis provided support for two groupings of problem behaviors - Externalizing and Internalizing.

In studying twin populations it is important to be able to generalize findings from the twin sample to the general population. The representativeness of the twin sample was studied in chapter 6. Maternal ratings of problem behaviors in twins, were compared with ratings of 2-3-year-old singletons whose mothers completed the CBCL/2-3. The twin sample consisted of 1281 twin pairs (407 MZ, 874 DZ), the singleton sample consisted of 420 children from the general population. Results indicated that the general level of problem behaviors in twins was broadly comparable to that in singletons. Five of the seven scales showed lower scores for DZ twins versus MZ twins and singletons. However, these differences were small. Standard deviations for 2 of the 7 scales were somewhat smaller for singletons than for twins. Higher means for boys were found for the total problem score, and the Aggressive and Overactive syndromes.

Chapter 7 reported, for the separate syndrome scales, the genetic analyses on the parental ratings of problem behaviors in their 3-year-old twins. The sample consisted of 218 MZ female, 189 MZ male, 233 DZ female, 252 DZ male, and 389 DZ opposite sex pairs. Both parents completed one CBCL/2-3 for each child. Model fit indices indicated that mothers and fathers assessed similar behaviors in their children. Genetic influences accounted on the average for 65% of the trait variance. Shared environmental influences accounted on the average for 12%, and non-shared environmental influences for 21% of the trait variance. Sex differences in genetic and environmental influences on problem behaviors were small. Evidence for sibling contrast effects was found for the Anxious and Overactive syndromes.

In chapter 8, multivariate genetic models were fitted to study patterns of problem behaviors in 3-year-old twins (446 MZ, and 912 DZ twin pairs). Fathers' and mothers' ratings of problem behaviors were obtained with the CBCL/2-3. A biometric model with two common genetic, one common shared environmental, and two common non-shared environmental factors fitted almost as well as the saturated unconstrained model for the genetic and environmental covariances. The common non-shared environmental factors produced externalizing/internalizing patterns of problem behaviors. One common genetic factor produced a clustering of the Oppositional, Withdrawn/Depressed, and Overactive syndromes with the Aggressive syndrome. The other common genetic factor produced a clustering of the Oppositional, Withdrawn/Depressed, and Overactive syndromes with the Anxious syndrome. A pattern of similar scores on all dimensions of problem behavior was most suggestive of the common shared environmental factor.

In the final chapter, chapter 9, results from the genetic analyses were discussed. Attention was paid to the interpretation of the findings, and issues concerning use and misuse of genetic findings were addressed. It was argued that heritabilities found in the present study imply that children show innate differences in liability to problem behavior,

and that misuse often is associated with misinterpretations of the quantitative genetic theory.

Samenvatting

Het primaire doel van dit proefschrift was om erfelijke- en omgevingsinvloeden op probleemgedrag bij kleuters en adolescenten te bestuderen. In hoofdstuk 1 werd beargumenteerd dat vanwege het continue karakter van de meeste psychiatrische condities bij kinderen de methodes van de kwantitatieve genetica kunnen worden toegepast, en dat stoornissen het beste gemeten kunnen worden als kwantitatieve variaties van gedrag in plaats van alles-of-niets categorieën. Het proefschrift werd opgedeeld in twee delen. Het eerste deel had betrekking op probleemgedrag bij kinderen en adolescenten. Dit deel bevatte erfelijke analyses op een steekproef van 11 t/m 15 jaar oude internationale adoptiekinderen. Het tweede deel betrof kleuters, en bevatte erfelijke analyses op probleemgedrag bij drie jaar oude tweelingen.

Deel 1. Een korte introductie in de methodes die toegepast werden om erfelijke invloeden op probleemgedrag bij kinderen en adolescenten te bestuderen werd gegeven in hoofdstuk 2. Deze introductie werd gevolgd door een overzicht van resultaten van erfelijke studies naar de meer algemene probleemgedragingen bij 4 t/m 18 jarigen. De kleine omvang van de steekproeven van de meeste studies in het overzicht, de verschillende manieren om probleemgedrag te meten, het gebruik van mogelijk incorrecte modellen, en het feit dat schattingen van erfelijke- en omgevingsinvloeden populatieafhankelijk zijn, maakte het moeilijk om duidelijke conclusies te trekken. In zover het mogelijk was algemene conclusies te trekken bleken erfelijke invloeden belangrijk te zijn voor de meeste probleemgedragingen. Gedeelde omgevings invloeden bleken invloed te hebben op antisociale gedragingen. De te lage DZ tweeling correlaties, die gevonden werden voor teruggetrokken gedrag en (hyper)activiteit, suggereerden dat voor dit soort gedrag het veelvuldig gebruikte model met additieve genetische-, gedeelde omgevings-, en ongedeelde omgevingsinvloeden mogelijk incorrect is.

De CBCL/4-18 (Child Behavior Checklist voor 4 t/m 18 jarigen) werd gebruikt om beoordelingen van probleemgedrag door ouders in de steekproef van internationale adoptiekinderen te krijgen. Het was de bedoeling om voor de erfelijke analyses recentelijk afgeleide Amerikaanse CBCL/4-18 syndromen te gebruiken. In hoofdstuk 3 werd daarom eerst de validiteit van de syndroomconstructen bestudeerd in de steekproef van internationale adoptiekinderen (N=2.148). Kruis-validatie werd verricht op een klinische steekproef (N=1.387). Resultaten ondersteunden de validiteit van de constructen. Maar de bijdrage van een aantal items aan de schalen van de syndroomconstructen was twijfelachtig in de steekproef van internationale adoptiekinderen. Deze items hadden een zeer geringe variantie, waren geen specifieke indicatoren van één construct, of verslechterden de betrouwbaarheid van de schaal.

In hoofdstuk 4 werden de Amerikaanse CBCL/4-18 syndromen, aangepast aan de Nederlandse steekproef van internationale adoptiekinderen, gebruikt om erfelijke- en omgevingsinvloeden op probleemgedrag te bestuderen. De steekproef (gemiddelde leeftijd 12.43 jaar) bevatte een groep van biologische broers/zussen (111 paar), een groep niet-biologische broers/zussen (221 paar), en een groep enigstkinderen (94). Ongedeelde omgevingsinvloeden waren het belangrijkst. Erfelijke invloeden waren substantieel for Externaliserende gedragingen, maar onbelangrijk voor Internaliserende gedragingen. De resultaten voor de CBCL totale probleemscore, Aandachts Problemen, en Externaliserende gedragingen waren in overeenstemming met resultaten van tweeling studies. Het gebrek aan erfelijke invloeden op Internalizerende gedragingen waren in contrast met resultaten van tweeling studies. Varianties voor enigstkinderen waren significant kleiner dan voor broers/zussen voor wat de Externaliserende groepering, Delinquent Gedrag, en

Agressief Gedrag betrof. Analyses gaven aan dat deze verschillen eerder werden veroorzaakt door factoren geassocieerd met het aantal kinderen in het gezin dan door actieve invloeden van broers/zussen op elkaar. Significante geslachtsverschillen werden gevonden voor 7 van de 10 schalen. De grotere varianties voor jongens voor de Externaliserende groepering en Agressief Gedrag werden veroorzaakt door erfelijke invloeden. Deel 2. Voor de erfelijke analyses op de steekproef van driejarige tweelingen werd de CBCL/2-3 (Child Behavior Checklist voor 2 t/m 3 jarige gebruikt) om ouderlijke beoordelingen van probleemgedrag te krijgen. In hoofdstuk 5 werd de Nederlandse factorstructuur van de CBCL/2-3 bestudeerd met drie verschillende steekproeven - een klinische steekproef, kinderen uit de algemene bevolking, en de tweeling steekproef. Exploratieve en confirmatieve factor analyses suggereerden een zeven-factor model voor alle drie de steekproeven. De syndromen werden als volgt benoemd: Oppositioneel, Teruggetrokken/Depressief, Agressief, Angstig, Overactief, Slaap Problemen, en Lichamelijke Klachten. Interne consistentie, test-hertest stabiliteit, en de overeenstemming tussen ouders was matig tot hoog voor de zeven factoren. Factor intercorrelaties en tweede-orde factor analyses suggereerden twee groeperingen van problemen - Externaliserend en Internaliserend.

Bij het bestuderen van tweelingpopulaties is het belangrijk om resultaten te kunnen generaliseren naar de algemene bevolking. De representativiteit van de tweeling steekproef werd bestudeerd in hoofdstuk 6. Beoordelingen van probleemgedrag door moeders in de tweeling steekproef werden vergeleken met beoordelingen van probleemgedrag van niet-tweelingen wier moeders ook de CBCL/2-3 hadden ingevuld. De tweeling steekproef bestond uit 1281 tweeling paren (407 MZ, 874 DZ), de niet-tweeling steekproef uit 420 kinderen uit de algemene bevolking. In het algemeen bleek het niveau van probleemgedrag bij tweelingen en niet-tweelingen hetzelfde. Vijf van de zeven schalen lieten wat lagere scores voor DZ tweelingen zien in vergelijking tot MZ tweelingen en niet-tweelingen. Deze verschillen waren echter klein. Voor twee van de zeven schalen waren de standaard afwijkingen iets kleiner voor niet-tweelingen dan voor tweelingen. Jongens hadden een wat hogere totale probleemscore, en scoorden wat hoger op de Agressief en Overactief syndromen.

In hoofdstuk 7 werden de erfelijke analyses op de afzonderlijke syndroomschalen gerapporteerd. De tweeling steekproef bestond uit 218 MZ meisjes, 189 MZ jongens, 233 DZ meisjes, 252 DZ jongens, en 389 DZ meisje/jongen paren. Beide ouders vulden één CBCL/2-3 voor elk kind in. Analyses gaven aan dat moeders en vaders hetzelfde gedrag beoordeelden. Erfelijke invloeden maakten 65% van de variantie in het gedrag zoals beoordeeld door beide ouders uit, gedeelde omgevingsinvloeden 12%, en ongedeelde invloeden 21%. Geslachtsverschillen in erfelijke- en omgevingsinvloeden waren klein. Bewijs voor broer/zus contrasteffecten werd gevonden voor de Angstig en Overactief syndromen.

In hoofdstuk 8 werden multivariate erfelijke modellen gepast om patronen van probleemgedrag in drie jarige tweelingen te bestuderen (446 MZ, en 912 DZ tweeling paren). Beoordelingen door vaders en moeders werden verkregen met de CBCL/2-3. Een biometrisch model met twee algemene genetische-, één algemene gedeelde omgevings-, en twee algemene ongedeelde omgevingsfactoren paste bijna even goed als het verzadigde niet-beperkte model voor de erfelijke- en omgevingsbijdrages aan de covarianties tussen de syndromen. De algemene ongedeelde omgevingsfactoren produceerden externaliserende en internaliserende patronen. Een algemene erfelijke factor produceerde een clustering van de Oppositioneel, Teruggetrokken/Depressief, en Overactief syndromen met het

Agressief syndroom. De andere algemene genetische factor produceerde een clustering van de Oppositioneel, Teruggetrokken/Depressief, en Overactief syndromen met het Angstig syndroom. Een patroon van gelijksoortige scores op alle dimensies van probleemgedrag was indicatief voor de algemene gedeelde omgevings factor.

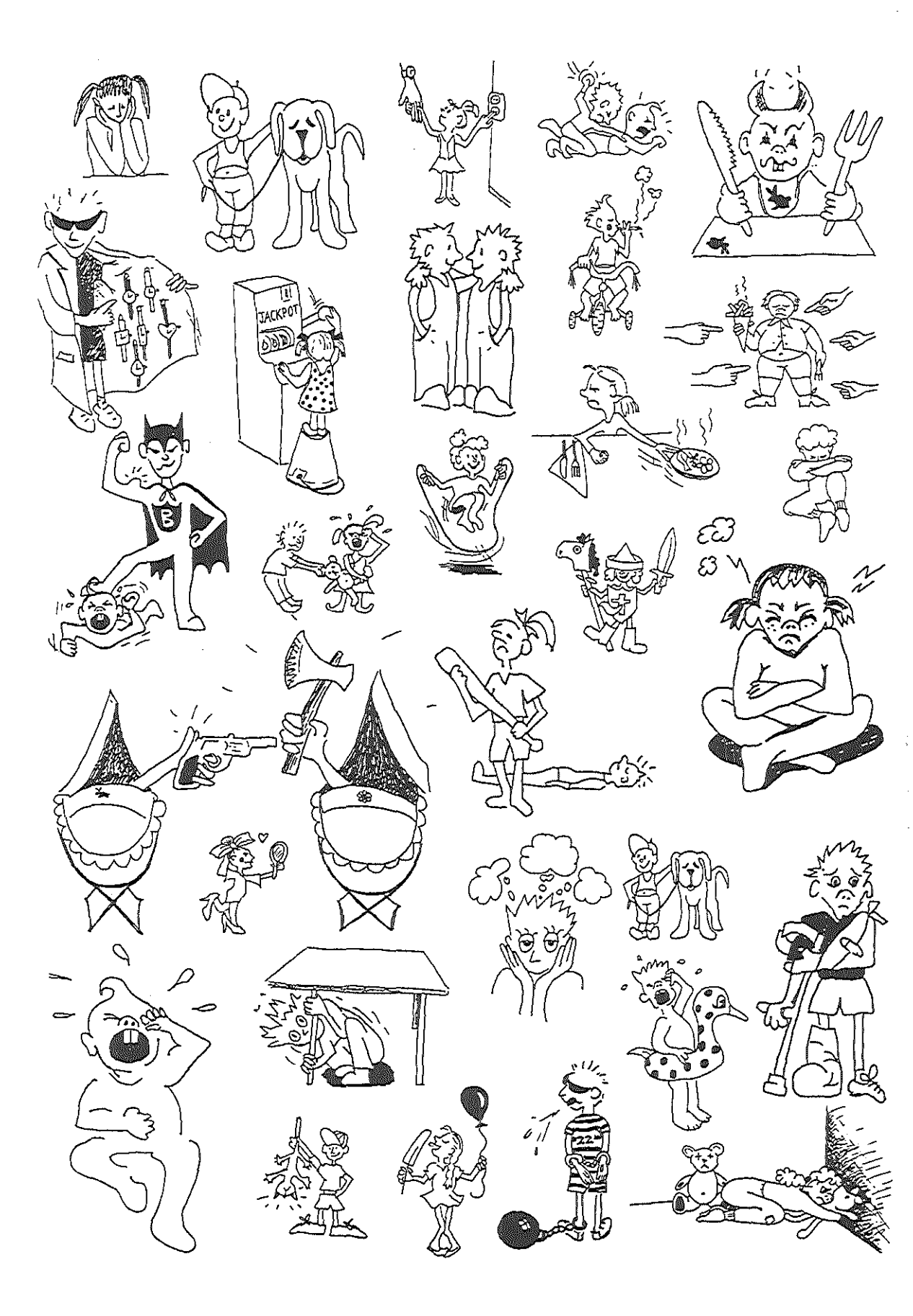
In het laatste hoofdstuk, hoofdstuk 9, werden de resultaten van de erfelijke analyses opgesomd, en aandacht werd besteed aan de implicatie en interpretatie van de resultaten. Er werd besproken dat een erfelijkheid zoals gevonden in deze studie impliceerd dat er tussen kinderen aangeboren verschillen in gevoeligheid voor probleemgedrag bestaan, en dat misbruik van erfelijke argumenten vaak geassocieerd is met misinterpretaties van de kwantitatief genetische theorie.

Curriculum vitae

Edwin van den Oord werd geboren op 19 april 1963 in Oegstgeest. In 1981 behaalde hij zijn V.W.O. diploma aan het Bonaventura college in Leiden. Hierna begon hij een studie lichamelijke opvoeding. In 1985 studeerde hij af aan de Academie voor Lichamelijke Opvoeding in Amsterdam. Dat zelfde jaar startte hij een studie psychologie aan de Universiteit van Amsterdam. In 1990 studeerde hij (cum laude) af als methodoloog bij de vakgroep methodenleer (hoofd van de vakgroep Prof. Dr. G.J. Mellenbergh). Van 1991 tot 1993 werkte hij als onderzoeker op een samenwerkingsproject tussen de afdeling kinder- en jeugdpsychiatrie van de Erasmus Universiteit Rotterdam (hoofd van de afdeling Prof. Dr. F.C. Verhulst), en de vakgroep Psychonomie van de Vrije Universiteit Amsterdam (hoofd van de vakgroep Prof. Dr. J.F. Orlebeke). Deze aanstelling resulteerde in dit proefschrift.

Correspondence to: Edwin van den Oord
Czaar Peterstraat 35^{2}
1018 NX Amsterdam
The Netherlands


22
JACKPOT


JACKPOT