Journal of Abnormal Psychology
1996, Vol. 105, No. 3, 349–357


0021-843X/96/$3.00

# A Genetic Study of Maternal and Paternal Ratings of Problem Behaviors in 3-Year-Old Twins

Edwin J. C. G. van den Oord
Sophia Children's Hospital and Free University

Frank C. Verhulst
Sophia Children's Hospital

Dorret I. Boomsma
Free University

Genetic and environmental influences on problem behaviors were studied in 3-year-old twins. Fathers' and mothers' ratings of problem behaviors in twins—236 monozygotic (MZ) girls, 210 MZ boys, 238 dizygotic (DZ) girls, 265 DZ boys, and 409 DZ opposite sex pairs—were obtained with the Child Behavior Checklist for Ages 2–3 (T. M. Achenbach, 1992). Twin correlations and results from a model fitting approach showed that genetic, shared environmental, and nonshared environmental influences accounted on average for about 64%, 9%, and 27% of the variance. Although shared environmental influences were small for most scales, they were important for Total Problems and somewhat larger for Externalizing than for Internalizing behaviors. Significant sex differences in genetic and environmental influences and evidence for sibling contrast effects were found for the Overactive scale.

During the 1960s and 1970s, many people emphasized the role of environmental influences in the etiology of children's problem behaviors (Rutter, 1991). In recent years, however, there has been an increased interest in the study of genetic factors (Rutter, Bolton, et al., 1990). These studies have led to a broader recognition that genetic factors may be involved in relatively rare child psychiatric conditions such as autism (Folstein & Rutter, 1977), tics (Pauls, Cohen, Heimbuch, Detlor, & Kidd, 1981), anorexia nervosa (Holland, Hall, Murray, Russel, & Crisp, 1984), and stuttering (Vandenberg, Singer, & Pauls, 1986), as well as in the more common varieties of children's problem behaviors like depression (Wierzbicki, 1987), hyperactivity (Goodman & Stevenson, 1989), delinquency (Rowe, 1983), and aggression (Ghodsian-Carpey & Baker, 1987; Plomin, Foch, & Rowe, 1981).

In contrast to the rare child psychiatric conditions that can be viewed as discrete categories that are either present or absent, the more common varieties of behavioral problems in children generally involve quantitative variations of behavior that most children display to some degree. 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 the methods of quantitative genetic theory should be used.

In quantitative genetics, twin or adoption data are used to disentangle genetic and environmental influences. The environmental influences are further divided into factors that have an impact on all children growing up in the same family and factors that affect children within a family differently. Parental rearing practices, illness or loss of a parent, or socioeconomic status are examples of possible shared environmental influences. Accidents, differential parental treatment, or peer group influences are examples of possible nonshared environmental influences. Disentangling genetic, shared environmental, and nonshared environmental influences may be scientifically and clinically useful because it offers a general framework for research efforts and clinical interventions.

Although quantitative genetics has already provided valuable contributions to the study of child psychiatric disorders, a number of limitations can be recognized. First, most of the studies focus on adolescents and children of school age, and very little is known about genetic influences on problem behaviors in preschool children. A number of authors showed that there is a substantial stability of problem behaviors (Richman, Stevenson, & Graham, 1982; Verhulst & Van der Ende, 1992) and indicated that early adjustment is an important predictor of the level of problem behavior at a later point in time. This argues for a better understanding of problem behaviors in young children, because an increased knowledge of the genetic and environmental determinants could help to optimize clinical interventions and prevent later maladjustment.

Second, in (genetic) research greater progress may 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 et al.,

Edwin J. C. G. van den Oord, Department of Child and Adolescent Psychiatry, Sophia Children's Hospital, Rotterdam, The Netherlands, and Department of Psychonomics, Free University, Amsterdam, The Netherlands; Frank C. Verhulst, Department of Child and Adolescent Psychiatry, Sophia Children's Hospital, Rotterdam, The Netherlands; Dorret I. Boomsma, Department of Psychonomics, Free University, Amsterdam, The Netherlands.

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

Correspondence concerning this article should be addressed to Frank C. Verhulst, Erasmus Universiteit Rotterdam, Sophia Kinderziekenhuis, Afdeling kinder-en jeugd psychiatrie, Dr. Molewaterplein 60, 3015 GJ Rotterdam, The Netherlands.

1986, p. 194). Narrowly defined syndromes such as hyperactivity, depression, and aggression may provide a better basis for detecting specific etiologies or predicting the outcome of specific treatments. However, in many genetic studies global measures of psychiatric dysfunctioning in children were used, making it difficult to specify to what behaviors the results apply.

A final set of limitations involves sample size and the methodology of quantitative genetic studies. In twin and adoption studies large samples are needed to reliably assess the magnitude of genetic and environmental influences (Martin, Eaves, Kearsey, & Davies, 1978). Sample sizes in most genetic studies of problem behaviors in children, however, are fairly small. The use of multiple raters for the assessment of problem behaviors also improves the reliability of estimates of genetic and environmental influences. A more important advantage is that with multiple raters, it is possible to apply corrections for rater bias (the tendency of an individual rater to overestimate or underestimate scores consistently as a consequence of stereotyping, response styles or different normative standards, etc.) and errors of measurement (Hewitt, Silberg, Neale, Eaves, & Erickson, 1992). Rater bias spuriously inflates estimates of the shared environment, and errors of measurement inflate estimates of the nonshared environment. Because assessments of a single rater were used in almost every genetic study of problem behavior in children, previous results were confounded with these methodological flaws.

The present article is one of the first genetic studies on problem behaviors in a large sample of preschool twins. For preschoolers parents are a primary source of information. Parental ratings of problem behaviors were obtained with the Child Behavior Checklist for Ages 2–3 (CBCL/2-3; Achenbach, 1992). The CBCL is a widely used rating scale that allows a distinction between a broad array of narrowly defined problem behaviors. To obtain multiple ratings, we asked mothers as well as fathers to complete the questionnaire.

## Method

### Participants

In The Netherlands, about 85% of the parents of newborns are paid a home visit by a commercial organization. During this visit parents of twins are asked to participate in the Netherlands Twin Register (NTR) kept by the Department of Psychonomics of the Free University in Amsterdam. Through this procedure, 40% to 50% of all multiple births in The Netherlands are registered (Boomsma, Orlebeke, & Van Baal, 1992).

Questionnaires were mailed to 1,792 parents of 3-year-old twins. Nonresponders were sent reminders and, when no response was obtained, contacted by phone. Completed questionnaires were returned for 1,377 twin pairs (77%).

For 403 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 the remainder of the twins zygosity information was obtained from a questionnaire completed by parents when the twins were about 2 years old. For 19 twin pairs zygosity could not be established. This procedure left a sample of 236 monozygotic (MZ) female, 210 MZ male, 238 dizygotic (DZ) female, 265 DZ male, and 409 opposite-sex pairs.

To assess the reliability of the questionnaire used to determine the twin's zygosity, we compared blood test results with the zygosity information from the questionnaire. For 356 same-sex twin pairs both blood test and questionnaire data were available. It could very well be that parents who were uncertain about their twin's zygosity were more likely to consent to a blood test. The proportion of correctly classified twins should probably therefore be viewed as the lower bound of the reliability of the questionnaire. The agreement between the two measures was 82%, implying that at most 7% ([.18% × 546]/1,358) of the twins were misclassified.

In an earlier article (Van den Oord, Koot, Boomsma, Verhulst, & Orlebeke, 1995), the demographic characteristics of the twin sample were presented, and twin-singleton differences in problem behaviors were studied. This study showed that the twin sample had fairly good epidemiological properties and that the level of problem behaviors in twins was broadly comparable to that of children from a community sample.

### Materials and Procedure

The CBCL/2-3 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 scales for the CBCL/2-3 were derived by Koot (1993) through applying exploratory factor analyses across three independent samples: children referred to mental health agencies; children from the general population; and the twin sample from the present study. The congruity found for the factor solutions was corroborated in a confirmatory factor analysis and made it possible to derive similar scales for the three samples. The scales were labeled Oppositional (Items 8, 13, 16, 29, 30, 33, 36, 44, 66, 69, 81, 82, 83, 85, 88, 96, 97; Cronbach's $\alpha = .91$); Withdrawn/Depressed (2, 23, 26, 43, 67, 70, 71, 77, 80, 90; $\alpha = .64$); Aggressive Behavior (14, 17, 18, 20, 35, 40, 42, 53, 91; $\alpha = .82$); Anxious (3, 4, 10, 21, 37, 68, 73, 87, 92; $\alpha = .83$); Overactive (5, 6, 11, 59, 62; $\alpha = .78$); Sleep Problems (22, 38, 48, 64, 74, 84, 94; $\alpha = .70$); and Somatic Problems (12, 52, 93; $\alpha = .59$). In addition, internalizing and externalizing groupings of problem behaviors were derived through second-order factor analyses (Koot, 1993). Internalizing consists of the items from the Anxious and Withdrawn/Depressed scales, and Externalizing consists of the items from the Aggression, Oppositional, and Overactive scales. Finally, an overall index of the number and severity of reported problems can be scored by summing all 99 items.

Similarities between the Dutch scales and those reported for American samples were studied by Koot (1993). By definition the scoring of Total Problems was identical. Correlations between Dutch and American versions were high for Externalizing problems ($r = .97$) and Internalizing problems (.90). For the separate scales these correlations were somewhat lower and ranged from .56 for Somatic Problems to 1.00 for Sleep Problems. Finally, in contrast to the Dutch version, there is no Overactive scale in the American version.

### Model

The mathematics and assumptions of the classical twin design have been discussed by several authors (Falconer, 1989; Plomin, DeFries, & McClearn, 1990). Individual differences in behavior are decomposed in genetic, shared environmental influences and nonshared environmental contributions. Furthermore, the model assumes that there is no assortative mating, correlations, or interactions between genes and environment and that MZ and DZ twins are raised in equal environments. Figure 1 presents a path diagram of the model. P1 and P2 represent the observed behavior of the first and second twin. The capital letter A refers to the additive genetic factor, C to the common (shared) environmental

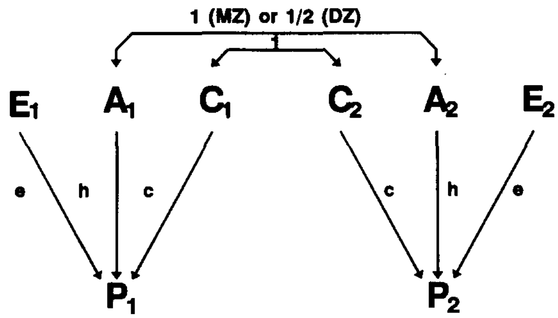


*Figure 1.* The basic genetic model. P = observed behavior, A = additive genetic factor, C = shared environmental factor, E = nonshared environmental factor, subscript 1 refers to the first twin, subscript 2 refers to the second twin, Parameter h = the additive genetic effect, c = the shared environmental effect, and e = the nonshared environmental effect. Genetic correlation equals 1.0 for monozygotic (MZ) twins and 0.5 for dizygotic (DZ) twins.

factor, and E to the nonshared environmental factor. Parameters h, c, and e represent the effect of A, C, and E, respectively.

The extent to which a trait is determined by genetic influences is called the *heritability* of the trait. The heritability equals the proportion genetic variance ($h^2$) of the total variance and indicates the extent to which differences in problem behaviors of children are caused by genetic influences.

Estimates of the heritability can be derived from differences in observed resemblance between MZ and DZ twin pairs. MZ twins are genetically identical; DZ twins share only a proportion of their genetic information. Therefore, a higher sibling resemblance for MZ twin pairs compared to DZ twin pairs suggests genetic influences. It can be shown (Falconer, 1989) that the heritability equals twice the difference between the MZ and DZ sibling correlations; in other words, $h^2 = 2(r_{MZ} - r_{DZ})$. The proportion of shared environmental variance can be estimated by $c^2 = 2r_{DZ} - r_{MZ}$, and the proportion nonshared environmental variance by $e^2 = 1 - r_{MZ}$.

## *Model Fitting*

In a structural modeling approach the observed variances and covariances for MZ and DZ twins are compared to the theoretical pattern implied by the model used for data analysis (Neale & Cardon, 1992). This approach offers a number of advantages compared to the traditional formulas that were presented above. First, more reliable estimates of genetic and environmental effects are obtained because the estimation procedures make optimal use of the available information. Second, significance tests can be performed to assess the fit of the model by comparing the observed versus implied variances and covariances. For instance, by specifying a model with no genetic influences, it is possible to test whether genetic influences are significant, and tests for sex differences can be performed by constraining genetic and environmental influences to be equal for boys and girls. Finally, the basic model in Figure 1 can be elaborated to account for more complex phenomena such as sibling interactions or nonadditive genetic effects.

To obtain the most parsimonious model, we performed a sequence of significance tests for each scale. First, genetic and environmental influences were estimated for boys and girls separately. Then estimates for genetic and environmental influences were constrained to each other when sex differences were not significant. Finally, nonsignificant estimates of genetic and environmental influences were fixed at zero. The model obtained after these tests was used to estimate the genetic and environmental influences.

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 (Heath, Neale, Hewitt, Eaves, & Fulker, 1989). For accurate significance testing, we performed logarithmic transformations to approximate normality (Muthén & Kaplan, 1985). For all significance tests, a probability level of .05 was applied.

## Results

Mother ratings were available for 99% and father ratings for 82% of the 1,377 twin pairs. Table 1 shows the twin correlations for the mother and father ratings and for the averaged parent rating.

For most scales, results for boys and girls and for mothers and fathers were surprisingly similar. In general, the twin correlations from Table 1 showed that genetic influences accounted on average for 62% of the variance, shared environmental influences for 9%, and nonshared environmental influences for 29%. For instance, the father ratings for same-sex MZ and DZ boys for Oppositional are a good example of the twin correlations that were found in the present study. The difference between MZ and DZ twin correlations suggested substantial genetic influences: $h^2 = 2x(.76 - .45) = .62$. The large MZ twin correlations indicated medium nonshared environmental effects: $e^2 = (1 - .76) = .24$. Relatively small shared environmental effect were suggested by the small differences between twice the DZ twin correlations minus the MZ twin correlation: $c^2 = (2x.45) - .76 = .14$.

The twin correlations for Total Problems differed from the other scales. For this scale the twin correlations were larger and the difference between the MZ versus DZ twin correlations was clearly smaller. This finding suggested that shared environmental influences are larger and genetic influences are smaller for Total Problems than for the other scales.

The Overactive scale showed substantial MZ twin correlations and near zero DZ twin correlations. This difference between MZ and DZ twin correlations implied genetic influences. Because DZ twins also share genes, a model with genetic, shared environmental and nonshared environmental influences would therefore have to yield larger DZ twin correlations. The low DZ twin correlations suggested that genetic dominance may contribute to the variance of this scale or that there could be sibling interactions.

Twin correlations for the mother ratings differed on average only .018 from the father ratings and indicated very similar estimates of genetic and environmental influences. The correlations computed for the average of mother and father ratings yielded small but consistently larger heritability estimates and smaller estimates of shared and nonshared environmental influences. This can be explained by the higher reliability of these ratings because they are based on two ratings of the same variable. This phenomenon, which resembles the effects of increasing the length of a psychometric test, results in larger amounts of variance attributable to the child's behavior and a reduced influence of errors of measurement and rater bias. Errors of measurement spuriously inflate estimates of the nonshared environment and rater bias inflates estimates of the shared environment. The use of the average rating therefore reduces the environmental component, thereby enlarging the relative im-

Table 1
*Twin Correlations for Mother (M), Father (F), and the Averaged (Avg.) Mother-Father Ratings*

| | MZ girls | | MZ boys | | DZ girls | | DZ boys | | DZ girls + boys (avg.) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | M | F | M | F | M | F | M | F | M | F |
| Total Problems | | | | | | | | | | |
| M | .88 | | .82 | | .65 | | .71 | | .65 | |
| F | .87 | .82 | .84 | .82 | .61 | .66 | .73 | .68 | .66 | .64 |
| Internalizing | | | | | | | | | | |
| M | .73 | | .64 | | .33 | | .43 | | .36 | |
| F | .80 | .74 | .69 | .65 | .35 | .39 | .47 | .41 | .35 | .36 |
| Externalizing | | | | | | | | | | |
| M | .81 | | .76 | | .50 | | .53 | | .54 | |
| F | .82 | .79 | .79 | .75 | .44 | .48 | .54 | .50 | .52 | .48 |
| Oppositional | | | | | | | | | | |
| M | .79 | | .72 | | .47 | | .51 | | .48 | |
| F | .79 | .76 | .76 | .72 | .40 | .45 | .53 | .48 | .44 | .42 |
| Depressed/Withdrawn | | | | | | | | | | |
| M | .68 | | .66 | | .32 | | .35 | | .44 | |
| F | .71 | .67 | .73 | .70 | .38 | .44 | .37 | .38 | .41 | .32 |
| Aggressive | | | | | | | | | | |
| M | .80 | | .80 | | .53 | | .51 | | .45 | |
| F | .83 | .77 | .81 | .76 | .49 | .42 | .49 | .44 | .45 | .42 |
| Anxious | | | | | | | | | | |
| M | .72 | | .53 | | .27 | | .38 | | .27 | |
| F | .77 | .70 | .62 | .58 | .31 | .34 | .41 | .35 | .28 | .31 |
| Overactive | | | | | | | | | | |
| M | .63 | | .40 | | .10 | | .10 | | .15 | |
| F | .61 | .54 | .42 | .35 | .04 | .06 | .12 | .13 | .08 | .09 |
| Sleep Problems | | | | | | | | | | |
| M | .70 | | .69 | | .49 | | .36 | | .31 | |
| F | .68 | .67 | .69 | .63 | .50 | .48 | .39 | .40 | .34 | .37 |
| Somatic Problems | | | | | | | | | | |
| M | .63 | | .73 | | .30 | | .24 | | .29 | |
| F | .63 | .54 | .73 | .59 | .29 | .19 | .26 | .28 | .27 | .25 |

*Note.* MZ = monozygotic; DZ = dizygotic.

portance of genetic influences (for an alternative approach, see the Appendix).

Because of its better psychometric properties, genetic models were fitted using the average of the mother and father rating. The use of the average rating assumes that both parents assess the same behavior and share a common understanding of the behavioral descriptions. The resemblance between the twin correlations of the mother and father ratings suggested that this may very well be true, and a joint analysis of maternal and paternal ratings of separate scale scales provided further support for this assumption (see the Appendix).

Table 2 shows for each of the scales the percentage of variance explained by genetic, shared, and nonshared environmental influences. These percentages were computed using the best fitting parsimonious model. The percentages from the model fitting approach confirmed the results that were derived from Table 1. Genetic, shared environmental, and nonshared environmental influences accounted on average for 65%, 9%, and 26% of the variance.

Except for Total Problems, genetic influences were clearly most important for all the scales and ranged from 63% for Somatic problems to 77% for Internalizing problems. Genetic influences tended to be slightly larger for problems associated with Internalizing behavior compared to Externalizing behavior.

For Total Problems, shared environmental influences accounted for 48% of the variance and were clearly larger than either genetic or nonshared environmental influences. For the other scales shared environmental influences were substantially lower or absent. This is a remarkable finding because most of the items in Total Problems are also in the other scales. Shared environmental influences were more important for Externalizing behavior compared to Internalizing behavior.

Nonshared environmental influences were significant for each scale and were equally important for Internalizing and Externalizing behavior. For the Overactive scale, sex differences in genetic and environmental influences were significant.

Except for the Overactive and Somatic Problems scales, the chi-square tests indicated a satisfactory fit. Chi-square tests often erroneously suggest that the model does not fit when scale scores depart too much from normality (Muthén & Kaplan, 1985). Therefore, the extreme nonnormality of the Somatic Problems scales seems the most likely cause for the model failure.

The Overactive scale, however, showed a normal distribution, and extreme nonnormality could not be the explanation for the

Table 2
*Percentages of Variance for Best-Fitting Model Explained by Genetic, Shared, and Nonshared Environmental Influences*

| | | Environment | | | | |
|---|---|---|---|---|---|---|
| Scale | Genetic | Shared | Nonshared | $\chi^2$ | *df* | *p* |
| Total Problems | 38 | 48 | 14 | 15.27 | 12 | .23 |
| Internalizing | 77 | — | 23 | 15.58 | 13 | .27 |
| Externalizing | 60 | 20 | 20 | 10.62 | 12 | .56 |
| Oppositional | 66 | 12 | 22 | 14.62 | 12 | .26 |
| Depressed/Withdrawn | 74 | — | 26 | 20.80 | 13 | .08 |
| Aggressive | 69 | 12 | 19 | 5.41 | 12 | .94 |
| Anxious | 72 | — | 28 | 17.15 | 13 | .19 |
| Overactive | | | | | 11 | |
| MZ girls | 66 | — | 34 | | | |
| MZ boys | 60 | — | 40 | | | |
| DZ girls | 70 | — | 30 | | | |
| DZ boys | 64 | — | 36 | 8.14 | 11 | .70 |
| Sleep Problems | 69 | — | 31 | 15.08 | 13 | .30 |
| Somatic Problems | 63 | — | 37 | 71.73 | 13 | .00 |

*Note.* MZ = monozygotic; DZ = dizygotic.

rejection of the model. In addition, the low DZ twin correlations that were found in Table 1 were also not consistent with the model that was used for data analysis. Too low DZ twin correlations can be caused by sibling contrast effects that occur when twins try to accentuate existing differences between them or when the behavior of one twin evokes a complementary reaction in the other (Carey, 1986; Eaves, 1976). Another explanation involves genetic dominance that refers to interactions between alleles at the same locus (Eaves, 1986).

To test these explanations, models with sibling contrast effects or genetic dominance were fitted. A significant better fit was obtained with both models. However, the chi-square indicated that the best fit was obtained with the model assuming sibling contrast effects. In addition, estimates obtained from fitting the model with genetic dominance indicated that these effects were even more important than additive genetic influences. From a theoretical point of view, large amounts of dominance are not plausible (Eaves, 1986). On the other hand, the parameter estimate for sibling contrast effect of −.143 was acceptable. Both fit indices and parameter estimates therefore indicated that the model with sibling effects should be preferred.

The finding of sibling contrast effects complicates the computations of the percentage genetic and environmental influences (exact formulas are given by Neale & Cardon, 1992, p. 208). These percentages depend on genetic correlation and are therefore different for MZ and DZ twins. Because of the sex differences the percentages are also different for opposite and same-sex twin pairs. However, these differences were very small and therefore were omitted from Table 2.

## Discussion

In the present article genetic and environmental influences on problem behaviors were studied in 3-year-old twins. Results indicated that genetic, shared environmental, and nonshared environmental influences accounted on average for about 64%, 9%, and 27% of the variance. Although shared environmental influences were small for most scales, they were important for Total Problems and somewhat larger for Externalizing than for Internalizing behaviors. Significant sex differences in genetic and environmental influences and evidence for sibling contrast effects were found for the Overactive scale.

There are only very few genetic studies of problem behaviors in preschool twins (e.g., Schmitz, Cherny, Fulker, & Mrazek, 1994), and a good comparison with other genetic studies was therefore not possible. Some of the scales in the present study like Oppositional or Activity resemble temperament characteristics like emotionality or attention span–activity. To the extent that our results can be compared with these genetic studies (Cohen, Dibble, & Grawe, 1977; Goldsmith & Gottesman, 1981; Matheny, Wilson, & Krantz, 1981; Neale & Stevenson, 1989; Plomin, 1986, p. 214; Wilson, Brown, & Matheny, 1971), heritabilities in the present study were similar or somewhat higher. Other scales like Internalizing, Externalizing, Aggressive Behavior, Withdrawn/Depressed, and Anxious have counterparts in problem behaviors found for older children. Results from these studies often show results that are rather inconsistent with each other (Edelbrock, Rende & Plomin, 1995; Ghodsian-Carpey & Baker, 1987; Hewitt et al., 1992; Plomin et al., 1981; Stevenson, Batten, & Cherner, 1992; Thapar & McGuffin, 1995; Van den Oord, Boomsma, & Verhulst, 1994; Wierzbicki, 1987). However, the results for Internalizing behaviors resemble findings from studies that showed the largest heritabilities (for an elaborate review of the relevant studies, see Van den Oord, Verhulst, & Boomsma, 1994).

If differences were found between the present study and studies of scales that have counterparts in older children or temperament characteristics in preschool children, the heritabilities from the present study tended to be somewhat larger. Partly this can be explained by the use of multiple raters that reduces the role of errors of measurement and rater bias and results in larger heritability estimates. In comparison to the present study, sample sizes were small in most of the genetic studies and sample fluctuations in these latter studies could have caused differences. However, genetic influences may also be more important for Internalizing behaviors in preschoolers compared to older children and for behavioral problems like the Oppositional scale compared to temperament characteristics like emotionality.

The ratio of genetic, shared environmental and nonshared environmental influences showed two large clusters. First, Internalizing, Depressed/Withdrawn, Anxious, Sleep Problems, and Somatic Problems showed a large genetic component, no shared environmental influences, and medium nonshared environmental influences. In contrast to Externalizing behavior, Oppositional and Aggressive behavior showed smaller genetic influences in favor of shared environmental influences. The large sample size could explain some of the uniformity that was obtained in the present study, because compared to other studies sampling errors are less influential in creating variation. However, it could also reflect that there is not yet much differentiation in the etiology of problem behaviors in young children.

Several twin studies of parental ratings of activity 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. These findings are not consistent with a model assuming additive genetic, shared environmental, and nonshared environmental influences. The large difference between the MZ and DZ twin correlations implies genetic influences. However, if genetic influences are important, the correlation for DZ twins, who share part of their genes, cannot be zero. In his book, Plomin (1986) explained the negative DZ twin correlations among others by the mechanism of sibling contrast effects, or as Buss and Plomin (1984, p. 119) 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 partner." The Overactive scale resembles the Activity scale reported by Plomin (1986). DZ twin correlations were too low in the Overactive scale as well. The model fitting approach used in the present study provided empirical evidence that the too low DZ twin correlations were caused by sibling contrast effects and made it possible to obtain an estimate of the size of these effects.

The evidence for significant shared environmental influences for Externalizing problems is consistent with other genetic as well as nongenetic research in the area of child psychopathology. McGuffin and Gottesman (1985) reviewed 6 twin studies concerning juvenile delinquency and crime and found a large shared environmental component. Evidence for shared environmental influences is not confined to results from twin studies. In three samples, Cadoret, Cain, and Crowe (1983) found a relation between psychopathology in adoptive parents or siblings and antisocial behaviors in adoptees. This finding rules out genetic explanations and clearly indicates a shared environmental component. Finally, a number of epidemiological studies have demonstrated the importance of aspects of the shared environment such as family discord, lack of affection, and poor supervision in conduct disturbance and antisocial behavior (Rutter, 1985). In contrast to most problem behaviors that were studied in the present article, Externalizing problems showed significant shared environmental influences. This indicated that even for 3-year-olds the shared environment has a demonstrable impact on Externalizing behaviors.

In the present study shared environmental influences were substantially larger for Total Problems compared to the other scales. This seems a little surprising because Total Problems comprises the items of these scales. Such a pattern, however, is consistently found in CBCL studies (Edelbrock et al., 1995; Van den Oord, Boomsma, & Verhulst, 1994). In addition, Plomin, DeFries, and Fulker (1988, pp. 183-184) reported substantial shared environmental influences for the CBCL Total Problem scale in a small sample of adopted and biological siblings. In part this finding could reflect the increased reliability of the total score that comprises much more items compared to the separate scales. However, there also seem to be other factors involved. In a genetic study the variance of scales is decomposed in genetic and environmental influences. The variance of the Total Problem scale consists of the variance of separate scales plus two times all the covariances between the scales. Thus, part of the explanation for the larger shared environmental influences for Total Problems compared to any of the separate scales could also be that shared environmental influences are especially important for causing covariances between scales. This hypothesis was tested and confirmed with a multivariate genetic analysis (Van den Oord, 1993). Thus, shared environmental effects do not seem to express themselves in a single problem but in multiple behavioral problems at the same time.

A meta-analysis by Achenbach, McConaughy, and Howell (1987) yielded correlations of about .6 between mothers' and fathers' reports of problem behaviors in the same child. The present study showed correlations of almost .7 and indicated a larger parental agreement. There are three possible explanations for the above average interparent correlations. It could be that parental agreement is larger for young children, because both parents observe most of their child's behavior in the same situation. This explanation is, however, not confirmed by other studies that reported interparent correlations for preschool children (Earls, 1989; Field & Greenberg, 1982). The higher interparent correlation could also indicate that CBCL scales are psychometrically superior compared to other assessment instruments because high reliability and few rater bias result in high interparent correlations. However, Koot (1993) studied the same scales in a sample of referred and nonreferred children and found the usual interparent correlations of about .6. The sample size in this study of 51 children was rather small and it is therefore not possible to draw firm conclusions on the basis of this finding. A final explanation involves the way questionnaires were completed. In the study by Koot (1993), a procedure was used that ensured that parents completed the questionnaires independently. In the present study this may not be the case, because parents completed the questionnaires at home. If this latter explanation for the larger interparent correlations is true, it cannot explain the higher heritabilities found in the present study. Instead it means that not all rater bias was eliminated from the analysis and that estimates of the shared environment were still somewhat too large.

For 18% of the sample, ratings for the mother only were available. This is probably because fathers were less cooperative, and they were absent in some of the families. If twins for whom only mothers responded differ from the other twins, selective nonresponse could limit the generalizability of the results when analyzing the father ratings or the average rating of both parents. Therefore, we compared both groups on a number of background variables (see Van den Oord et al., 1995, for a more elaborate discussion of these background variables). For each parent, information was available on age, education, occupational level, and parity. In addition, these families were compared on variables that apply to the twins themselves, such as number of hours spend in day care, the prevalence of illness or handicaps, hospital admission, and whether the child has been anaesthetized. Significant differences were found only for the age of both parents and the educational level of the father.

In the mother-only families, both parents were older and the educational level of the father was lower. The twin correlations in both groups were also studied. In the mother-only families, the twin correlations were about .05 higher than the twin correlations in the group for whom ratings of both parents were available (the differences were significant for Total Problems, Externalizing grouping, Oppositional Behavior, and Sleep Problems). This implies that 5% more shared environmental influences in the former group compared to the latter. However, this group constitutes only 18% of the total sample; the inclusion of the mother-only group would increase shared environmental influences only by $.18 \times 5\% = .9\%$. Thus, even if it is assumed that all families in the mother-only group differ from the other families on relevant characteristics, the impact on the results for the whole sample could not make a large difference.

The present study showed that genetic influences are important for problem behaviors in preschool children, and perhaps even more important compared to problem behaviors in older children. The average heritability of .64 indicated that 64% of the differences between childrens' problem behaviors could be explained by genetic factors. This suggests that children with behavior problems are likely to show an innate vulnerability. As emphasized by many authors, a high heritability does not mean that the behavior of concern is unchangeable (e.g., Rutter, 1991; Vandenberg & Crowe, 1989). The finding of genetic effects implies hereditary propensities, not predestination (Plomin & Daniels, 1986).

The heritability is an index for average differences among individuals in a population. This means that at an individual level there can be total environmental etiology for some children and total genetic etiology for others. A related issue is that genetic and environmental etiologies that completely explain a disorder for a specific group of individuals account for a negligible amount of variance in the population as a whole and could thus remain undetected in analyses of total variations in the population (Plomin, Rende, & Rutter, 1991).

To apply the findings from the present study to psychiatric conditions in clinically referred children, 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 more common varieties of emotional and conduct disturbances (Rutter, Macdonald, et al., 1990), genetic and environmental etiologies may also be different for clinical and nonclinical populations. For instance, clinically referred depressed individuals do not need to be affected by the same genes or environments as "mood" problems in the general population.

## References


Achenbach, T. M. (1991). *Integrative Guide for the 1991 CBCL/4-18, YSR, and TRF Profiles.* Burlington: University of Vermont, Department of Psychiatry.

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

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,* 213-232.

Boomsma, D. I., Orlebeke, J. F., & Van Baal, G. C. M. (1992). The Dutch twin register: Growth data on weight and height. *Behavior Genetics, 22,* 247-251.

Buss, A. H., & Plomin, R. (1984). *Temperament: Early developing personality traits.* Hillsdale, NJ: Erlbaum.

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.

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics, 16,* 319-341.

Cohen, D. J., Dibble, E., & Grawe, J. M. (1977). Fathers' and mothers' perceptions of children's personality. *Archives of General Psychiatry, 34,* 480-487.

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. J. (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. (1995). Genetic and environmental effects on competence and problem behaviors in childhood and early adolescence. *Journal of Child Psychology and Psychiatry, 36,* 775-785.

Falconer, D. S. (1989). *Introduction to quantitative genetics.* Essex, England: 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.

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

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., & Gottesman, I. I. (1981). Origins of variation in behavioral style: A longitudinal study of temperament in young twins. *Child Development, 52,* 91-103.

Goodman, R., & Stevenson, J. (1989). A twin study of hyperactivity-II. The aetiological role of genes, family relationships and perinatal adversity. *Journal of Child Psychology and Psychiatry, 30,* 691-709.

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.

Hewitt, J. K., Silberg, J. L., Neale, M. C., Eaves, L. J., & Erickson, M. (1992). The analysis of parental ratings of children's behavior using LISREL. *Behavior Genetics, 22,* 293-318.

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.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7. A guide to the program and applications.* Chicago: SPSS, Inc.

Koot, H. M. (1993). *Problem behavior in Dutch preschoolers.* Unpublished doctoral dissertation, Erasmus University Rotterdam, Rotterdam, The Netherlands.

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

Matheny, A. P., Wilson, R. S., & Krantz, J. Z. (1981). Behavioral contrasts in twinship: Stability and patterns of differences in childhood. *Child Development, 52*, 579–588.

McGuffin, P., & Gottesman, I. I. (1985). Genetic influences on normal and abnormal development. In M. Rutter & L. Hersov (Eds.), *Child and adolescent psychiatry: Modern approaches* (2nd ed., pp. 17–33). Oxford, England: Blackwell Scientific.

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, The Netherlands: Kluwer Academic.

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.

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.

Plomin, R. (1986). *Development, genetics and psychology.* Hillsdale, NJ: Erlbaum.

Plomin, R., & Daniels, D. (1986). Genetics and shyness. In W.H. Jones, J.M. Cheek, & S.R. Briggs (Eds.), *Shyness: Perspectives on research and treatment* (pp. 63–80). New York: Plenum Press.

Plomin, R., DeFries, J. C., & Fulker, D. W. (1988). *Nature and nurture during infancy and early childhood.* New York: Cambridge University Press.

Plomin, R., DeFries, J. C., & McClearn, G. E. (1990). *Behavioral genetics: A primer.* San Francisco: Freeman.

Plomin, R., Foch, T. T., & Rowe, D. C. (1981). Bobo clown aggression 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 & S. Miller (Eds.), *Handbook of developmental psychopathology* (pp. 119–133). New York: Plenum.

Plomin, R., Rende, R. D., & Rutter, M. L. (1991). Quantitative genetics and developmental psychopathology. In D. Cicchetti & S. Toth (Eds.), *Rochester symposium on developmental psychopathology: Vol. 2. Internalizing and externalizing expressions of dysfunction* (pp. 155–202). Hillsdale, NJ: Erlbaum.

Richman, N., Stevenson, J., & Graham, P. J. (1982). *Preschool to school: A behavioral study.* New York: Academic Press.

Rowe, D. C. (1983). Biometrical genetic models of self-reported delinquent behavior: A twin study. *Behavior Genetics, 13*, 473–490.

Rutter, M. (1985). Family and school influence on behavioural development. *Journal of Child Psychology and Psychiatry, 26*, 349–368.

Rutter, M. (1991). Nature, nurture and psychopathology: A new look at an old topic. *Development and Psychopathology, 3*, 125–136.

Rutter, M., Bolton, P., Harrington, R., Le Couteur, A., Macdonald, H., & Simonoff, E. (1990). Genetic factors in child psychiatric disorders—I. A review of research strategies. *Journal of Child Psychology and Psychiatry, 31*, 3–37.

Rutter, M., Macdonald, H., Le Couteur, A., Harrington, R., Bolton, P., & Bailey, A. (1990). Genetic factors in child psychiatric disorders—II. Empirical findings. *Journal of Child Psychology and Psychiatry, 31*, 39–83.

Schmitz, S., Cherny, S. S., Fulker, D. W., & Mrazek, D. A. (1994). Genetic and environmental influences on early childhood behavior. *Behavior Genetics, 24*, 25–34.

Stevenson, J., Batten, N., & Cherner, M. (1992). Fears and fearfulness in children and adolescents: A genetic analysis of twin data. *Journal of Child Psychology and Psychiatry, 33*, 977–986.

Thapar, A., & McGuffin, P. (1995). Are anxiety symptoms in childhood heritable? *Journal of Child Psychology and Psychiatry, 36*, 439–447.

Torgersen, A. M. (1982). Influence of genetic factors on temperament development in early childhood. In R. Porter & G. M. Collins (Eds.), *Temperamental Differences in Infants and Young Children* (Ciba Foundation Symposium 89, pp. 141–154). London: Pitman.

Vandenberg, S. G., & Crowe, L. (1989). Genetic factors in childhood psychopathology. In B. B. Lahey & 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. (1993). *A genetic study of problem behaviors in children.* Unpublished doctoral dissertation, Erasmus University Rotterdam, Rotterdam, The Netherlands.

Van den Oord, E. J. C. G., Boomsma, D. I., & Verhulst, F. C. (1994). A study of problem behaviors in 10- to 15-year-old biologically related and unrelated international adoptees. *Behavior Genetics, 24*, 193–205.

Van den Oord, E. J. C. G., Koot, H. M., Boomsma, D. I., Verhulst, F. C., & Orlebeke, J. F. (1995). A twin-singleton comparison of problem behaviour in 2–3-year-olds. *Journal of Child Psychology and Psychiatry, 36*, 449–458.

Van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (1994). A review of behavior genetic studies of problem behavior in children and adolescents. In *Genetic epidemiology of twins and twinning* (pp. 130–151). Amsterdam: Free University.

Verhulst, F. C., & Van der Ende, J. (1992). Six-year developmental course of internalizing and externalizing problem behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*, 924–931.

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.

Wilson, R. S., Brown, A. M., & Matheny, A. P. (1971). Emergence and persistence of behavioral differences in twins. *Child Development, 42*, 1381–1398.

# Appendix

## Multivariate Analyses of Parental Ratings

Hewitt et al. (1992) discussed the application of three classes of models for the joint analysis of mothers' and fathers' ratings of twins. The first model, 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. In contrast to the biometric models, the bias model 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 the tendency of an individual rater to overestimate or underestimate scores consistently compared to the mean of the raters (rater bias) and unreliability of the assessment instrument. The final model, the psychometric model, 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 that is unique to each rater.

These three models were fitted to the twin data from the present article (Van den Oord, 1993). Models fit indexes showed that the biometric model never fit much better than either the psychometric model or the bias model and indicated that mothers and fathers assessed similar behaviors in their children. This finding indicates that ratings of mothers and fathers of problem behaviors in their 3-year-olds are for the major part expressions of the same underlying trait and justifies averaging the ratings to obtain an estimate of this trait.

In addition, results obtained from the best fitting parsimonious model resembled those obtained from the genetic models using the average rating of both parents that were presented in this article. Thus, genetic influences were largest and accounted on average for about 65% of the variance. Shared environmental influences were least important and accounted on average for 12%. Nonshared environmental influences accounted for 21% of the variance. Total Problems showed large shared enviromental influences (46%); for Overactive, evidence was found for sex differences and sibling contrast effect.

Other, minor, differences could be explained by the larger power of the significance tests. For instance, because of the larger power the models did not fit as nicely as in the analyses with the average rating of both parents that were performed in the present study. However, results from these complex analyses were in general similar to the much more plain analyses on the averaged parent rating, and so these latter analyses were presented in this article.

Received February 28, 1994
Revision received April 20, 1995
Accepted October 3, 1995 ■