



Acta Genet Med Gemellol 41: 53-63 (1992)
© 1992 by The Mendel Institute, Rome

Received 30 April 1992
Final 28 May 1992

Childhood Behavior Problems: A Comparison of Twin and Non-twin Samples

J.S. Gau¹, J.L. Silberg², M.T. Erickson¹, J.K. Hewitt²

¹Department of Psychology, Virginia Commonwealth University, Richmond; ²Department of Human Genetics, Medical College of Virginia, Richmond, USA

Abstract. This study compares standardized measures of childhood behavior problems in a community-based twin sample with those for normative samples from the general population. Maternal parent ratings on the Child Behavior Checklist (CBCL) for 1824 twins were compared with the CBCL normative sample. The results indicated that twins showed small but consistently higher levels of problem behaviors. These elevations were significant for older children on both internalizing and externalizing behaviors; for younger children the elevations were significant for externalizing but not internalizing behaviors.

Key words: Child behavior, Twin children, Problem behavior

INTRODUCTION

Twin studies are increasingly being used to assess genetic factors in the development of childhood behavior disorders [5]. In studying twin populations it is important to be able to generalize findings from the twin sample to the general population [4]. However, the increased occurrence of some types of perinatal insult in twins as compared to singletons [8] raises questions about the equivalence of twin and general population samples. This study will assess one hypothesis aimed at evaluating this equivalence, with specific attention to the issue of perinatal injury and brain damage.

Well documented differences between pregnancies in single and multiple births, including lower birth weight and shorter length of gestation in twins, are known to increase mortality rates in twin pregnancies [3]. Higher frequencies of abnormalities in pregnancy and delivery have also been observed in twins as compared to singletons. For example, the rate of assisted breech deliveries in twins is about 16.96% for the first born twin and 32.52% for the second twin, as compared to 1.72% for singletons. Use of for-

ceps is similarly elevated for twins, with 18.70% of twin deliveries requiring forceps compared to 11.38% of singleton deliveries [7]. Rate of Cesarean Sections is also higher in twins (7.65%) than in singletons (5.92%) [7]. Thus, overall risk of perinatal injury is greater for twins.

Many of the high risk pre- and perinatal situations involve complications that can lead to fetal brain injury. These injuries have been implicated in elevations in rate and severity of children's behavior disorders, particularly externalizing behaviors, such as, hyperactivity and conduct disorder [13,2]. While this relationship is strong in cases of severe brain injury, the data for low level injuries, often called "minimal brain damage", are less conclusive. Pasamanick and Knobloch [9] suggested a "continuum of reproductive casualty", in which the severity of brain injury affected the severity of resulting problems. Severe brain damage along this continuum would produce notable neurological disorders, such as, cerebral palsy or epilepsy; less severe injury would lead to disruptions in behavioral development and associated increases in problematic behaviors.

Minimal brain damage is difficult to assess directly, however, because low level injuries often do not appear in neurological or psychological examinations [10]. One approach to assessing the effects of minimal brain damage on children's externalizing behavior is to compare groups, such as twins, that have a higher risk for pre- and perinatal brain injury with groups of nontwins that have an average risk for such injuries. Though brain injuries may not be physically identifiable, significant elevations in rates of behavioral problems associated with minimal brain damage would be expected for twins as compared to non-twins.

The current study provides a general, albeit simplistic, comparison between twins and non-twins, addressing the minimal brain damage issue. In order to compare the twin population with a non-twin population, our sample is compared to Achenbach's normative sample for the Child Behavior Checklist (CBCL), with means, standard deviations, and standard errors taken from the *Manual for the Child Behavior Checklist* [1]. In making this comparison, we hypothesized that scores for externalizing behaviors in particular would be slightly higher in our twin sample than in the Achenbach non-twin normative sample, due to the possible contribution of perinatal injury.

METHODS

Subjects

The twin sample consisted of 3716 Child Behavior Checklists, completed by mothers or female guardians of twins registered with the Virginia Twin Register, from ages 6 through 16. The questionnaires were mailed from 27 July 1989 to 31 August 1990, reaching 4204 mothers/female guardians (2 CBCLs were completed by each parent, one for each twin); 1858 mothers/female guardians completed and returned both CBCLs, making a total of 3716 CBCLs received and a response rate of 44%.

Several alterations from the response rate sample occurred in producing the twin sample that was actually used in this analysis. The first involves the method of calculating the age of each subject. In the final twin sample, age was calculated according to

Table 1 - Subjects' Race in Each Sample

| | CBCL sample | Twin sample |
|------------------|-------------|-------------|
| Caucasian | 80.5% | 81.0% |
| African-American | 18.2 | 15.1 |
| Other/Missing | 1.3 | 3.9 |

the date each CBCL was received minus date of birth. In calculating response rate, however, the date each CBCL was sent had to be substituted for the date each CBCL was received in the age calculation formula. This resulted in a net loss of 8 CBCLs from the response rate sample to the final sample, leaving 3708 CBCLs. In addition, 16 CBCLs that were included in the response rate calculation were ineligible for the final sample because they were received after the cutoff date for the current analysis. This reduced the sample to 3692 CBCLs. Finally, 6 CBCLs in the final sample lacked birth date information that was available in the response rate sample, thus they were not included in the analysis. These alterations left a total of 3686 CBCLs in the twin sample.

Two additional modifications were made in order to equate the twin sample with the non-twin sample. First, one twin from each pair was randomly selected, thus including only one child from each family to match the CBCL sample, leaving 1843 CBCLs. In addition, 19 subjects whose scores included more than 8 missing answers were deleted from the sample, as recommended by the scoring rules in the CBCL Manual [1]. The final sample thus consisted of 1824 CBCLs.

Age of the twins who were rated had a mean of 11.0 years and a standard deviation of 2.9 years. Female twins comprised 49.7% of the sample; 50.3% were male. Data on race of subjects in the twin sample was similar to that of the CBCL sample, with the twin sample having a slightly lower rate of non-Caucasian subjects (see Table 1). The CBCL Manual [1] reports that differences between races were not significant when socioeconomic status was controlled. Precise socioeconomic status information was not available for the twin sample, however, thus a direct comparison of the two samples along these variables was not possible. Nevertheless, non-Caucasian subjects tend to obtain higher ratings on the CBCL, so a decrease in the number of Caucasian subjects in the twin sample would likely increase the mean scores of the twin sample, thus providing further support for our hypothesis.

The CBCL normative sample, as described in the CBCL Manual [1], was selected randomly from homes located in Maryland, Northern Virginia and Washington, DC. Of the 1752 parents contacted by interviewers for the normative sample, 82.3% completed the questionnaire. This response rate is notably higher than that found in the twin sample.

Two factors may help explain the difference in response rates between the CBCL sample and the twin sample. The first involves the method of data collection. The twin sample was based on CBCLs obtained from a list of eligible twins in the state of Virginia. In contrast, the CBCL normative sample was based on CBCLs obtained through door-to-door interviews. The interviewers were instructed to attempt an interview at a target home, selecting a child between 4 and 16 years old in the household as

the target of the interview (using random number tables to make the selection once the children in a household had been enumerated). Interviewers continued gathering data until 50 subjects were obtained at each age and gender category. The 82.3% response rate was based on the number of parents contacted, rather than the total number of target homes selected. This procedure clearly raises the response rate for the CBCL sample, and, unlike the twin sample, does not require a representative sample of previously identified children. Thus, the expected participation rate of the twin study would be lower than the participation rate of the CBCL study.

The second explanation of the difference in response rates is the twin study's reliance upon parents to return questionnaires through the mail, a process that typically yields lower participation rates than an interview survey like the CBCL sample. Thus, although the twin study response rate was lower than that of the CBCL sample, the 44% response rate for the twin sample is consistent with that typically observed in studies involving the Virginia Twin Registry [11,6].

Questionnaires in the CBCL sample were completed by mothers whenever possible (83.1%), with a small percentage of fathers participating (13.5%). The remaining 3.4% of participants completing the questionnaires were classified as "other", which included relatives, foster parents or other guardians. Data for the twin sample were taken from checklists completed by the closest female guardian only (ie, mother, stepmother, grandmother, adopted mother, foster mother, or other female guardian). The distributions of total problem scores rated by mothers and fathers were not equivalent, with mothers' ratings having a mean of 24.2 (sd = 19.2) and fathers' ratings having a mean of 19.9 (sd = 16.7) ($t = 8.0$, $p < 0.01$).

Although including more fathers' ratings would decrease the twins' total behavior problem scores, an estimate of the reduction in mothers' ratings can be calculated, using the ratio of mothers to fathers in the CBCL sample. This would produce an estimated mean score of $[(0.87 \times 24.2) + (0.17 \times 19.9)] = 23.5$. This estimate, when compared to the actual mean mother rating of 24.2 does not indicate a reduction great enough to explain the elevation of twin scores compared to singleton scores.

A final difference between selection of subjects in the CBCL sample and subject selection in the twin sample involves the inclusion of children with clinical diagnoses. While the twin sample did not screen for clinically treated behavior problems, the CBCL sample excluded children who had received mental health services during the previous year. The CBCL Manual [1] does not report the percentage of children excluded by this criterion.

Instrument

The Child Behavior Checklist (Parent Report Form) was used to assess twins' behavior in both samples. This form provides a standardized method for assessing children's behavior as rated by parents.

The behavior assessment section of the CBCL consists of 118 questions that describe a wide range of children's problems, relevant to mental health referrals. The items are rated on a three-step scale; parents are asked to rate the occurrence of each behavior item currently or within the past six months, circling 0 if the item is not true, 1 if the item is somewhat or sometimes true, and 2 if the item is often or very true for the child.

Several scores are generated from the CBCL items. An overall score for behavior problems is created by totaling the scores on each item. A first order factor analysis on the items revealed individual subscales which corresponded to specific syndromes observed in children's behavior problems. These subscales include symptoms related to somatic complaints, schizoid behaviors, withdrawal, obsessions and compulsions, depression, anxiety, hyperactivity, aggression, and delinquency. Items falling under each subscale are totaled, and converted into standardized scale scores.

A second order factor analysis produced two broader categories representing internalizing and externalizing behaviors. Most of the subscales above load on one of the two broader internalizing/externalizing scales. Items loading on these factors are also totaled to produce standardized scores.

The current edition of the CBCL distinguishes for each subscale between boys and girls, as well as between older (12-16 years) and younger (6-11 years) children. Thus, each subscale is created within one of four groups: boys aged 6-11, boys aged 12-16, girls aged 6-11 or girls aged 12-16. Items loading on a particular scale for one group may not load on the corresponding scale for another age/gender group, however, thereby complicating comparisons between groups. As a result, comparisons between the two samples in this study are made only between the four age/gender group.

The CBCL Manual [1] reports reliability for the individual items, scale scores, and agreement between mothers' and fathers' scale scores. For the item scores, test-retest reliabilities were in the 0.90's, including interparent agreement and interinterviewer agreement. For scale scores and total problem scores, the median test-retest reliability for mother's ratings was 0.89.

The validity of the CBCL is also reported in the Manual. Of the 118 behavior items, 116 were significantly ($p < 0.01$) associated with an independent rating of clinical status, suggesting excellent concurrent validity. Correlations between CBCL total behavior problem scores and similar scores from other popular parent rating forms ranged from about 0.75 for young girls to about 0.91 for young boys. These correlations provide evidence for construct validity in the CBCL.

RESULTS

Using t-tests for differences between mean scores, scales with significant differences were identified, as shown in Tables 2-5. Unequal n's at each age in the twin sample required the calculation of the unweighted means for each age group in order to compare means with the normative sample, which had 50 subjects at each age in each of the four age groups (boys aged 6-11, boys aged 12-16, girls aged 6-11, and girls aged 12-16). These calculations thus ensure that mean differences between the twin sample and the normative sample are not due to differences in subjects' age.

Overall, the distribution for the two samples appears similar, with ratings for twins slightly higher than those for the normative sample. The Hyperactive, Aggressive and Delinquent scales were significantly higher for the twin sample for the CBCL sample across all age and gender groups except young girls. The Externalizing scale and Total Problem scale were significantly elevated in all twin groups. The Internalizing scale was elevated for the older twin groups only.

Table 2 - Mean scale scores for boys aged 6-11

| | CBCL | | Twin sample (Female raters only) | | | |
|-----------------------------|--------------------|-----------|----------------------------------|-----------|------------------|-----------|
| | Nonclinical sample | | Weighted means | | Unweighted means | |
| | Raw score | SD of raw | Raw score | SD of raw | Raw score | SE of raw |
| | N = 300 | | N = 512 | | N = 512 | |
| Internalizing scales | | | | | | |
| Schizoid or Anxious* | 1.3 | 1.4 | 1.4 | 1.6 | 1.6 | 0.1 |
| Depressed | 3.2 | 3.4 | 3.6 | 3.8 | 3.7 | 0.2 |
| Uncommunicative* | 2.0 | 1.9 | 2.2 | 2.2 | 2.3 | 0.1 |
| Obsessive-Compulsive | 2.9 | 2.8 | 2.9 | 3.0 | 2.9 | 0.1 |
| Somatic | 0.8 | 1.3 | 1.0 | 1.6 | 1.0 | 0.1 |
| Social withdrawal | 1.7 | 1.8 | 1.8 | 2.1 | 1.8 | 0.1 |
| Externalizing scales | | | | | | |
| Hyperactive* | 3.2 | 2.9 | 3.9 | 3.5 | 3.9 | 0.2 |
| Aggressive* | 7.3 | 5.7 | 8.7 | 7.1 | 8.3 | 0.3 |
| Delinquent** | 1.0 | 1.7 | 1.7 | 2.3 | 1.6 | 0.1 |
| Internalizing | 8.4 | 6.7 | 9.2 | 8.1 | 9.3 | 0.4 |
| Externalizing** | 10.8 | 8.2 | 13.1 | 10.4 | 12.8 | 0.5 |
| Total** | 21.7 | 15.0 | 25.6 | 18.7 | 25.5 | 0.8 |

T-test comparing CBCL sample means with twin sample unweighted means: *p < 0.05 **p < 0.01

Table 3 - Mean scale scores for boys aged 12-16

| | CBCL | | Twin sample (Female raters only) | | | |
|-----------------------------|--------------------|-----------|----------------------------------|-----------|------------------|-----------|
| | Nonclinical sample | | Weighted means | | Unweighted means | |
| | Raw score | SD of raw | Raw score | SD of raw | Raw score | SE of raw |
| | N = 250 | | N = 395 | | N = 395 | |
| Internalizing scales | | | | | | |
| Somatic ** | 1.4 | 2.0 | 2.3 | 2.9 | 2.3 | 0.1 |
| Schizoid ** | 1.1 | 1.5 | 1.5 | 2.0 | 1.5 | 0.1 |
| Uncommunicative ** | 3.2 | 3.6 | 4.1 | 4.1 | 4.1 | 0.2 |
| Immature ** | 0.9 | 1.3 | 1.6 | 2.1 | 1.5 | 0.1 |
| Obsessive-Compulsive | 1.7 | 1.9 | 2.0 | 2.2 | 1.9 | 0.1 |
| Hostile withdrawal ** | 1.8 | 2.5 | 3.0 | 3.6 | 2.8 | 0.2 |
| Externalizing scales | | | | | | |
| Delinquent ** | 1.2 | 2.0 | 1.8 | 2.7 | 1.8 | 0.1 |
| Aggressive ** | 5.7 | 5.9 | 7.2 | 6.9 | 7.3 | 0.4 |
| Hyperactive ** | 3.0 | 2.9 | 4.0 | 3.6 | 3.9 | 0.2 |
| Internalizing ** | 7.4 | 7.4 | 10.3 | 9.5 | 10.0 | 0.4 |
| Externalizing ** | 8.4 | 8.4 | 11.1 | 10.2 | 11.1 | 0.5 |
| Total ** | 17.5 | 15.6 | 23.9 | 20.2 | 23.7 | 1.0 |

T-test comparing CBCL sample means with twin sample unweighted means: * p < 0.05 ** p < 0.01

Table 4 - Mean scale scores for girls aged 6-11

| | CBCL | | Twin sample (Female raters only) | | | |
|-----------------------------|--------------------|-----------|----------------------------------|-----------|------------------|-----------|
| | Nonclinical sample | | Weighted means | | Unweighted means | |
| | Raw score | SD of raw | Raw score | SD of raw | Raw score | SE of raw |
| | N = 300 | | N = 541 | | N = 541 | |
| Internalizing scales | | | | | | |
| Depressed | 4.2 | 3.7 | 5.3 | 4.7 | 4.6 | 0.2 |
| Social withdrawal** | 1.8 | 1.9 | 2.6 | 2.6 | 2.3 | 0.1 |
| Somatic** | 1.7 | 2.0 | 2.4 | 2.6 | 2.2 | 0.1 |
| Schizoid-Obsessive | 0.7 | 1.1 | 0.7 | 1.4 | 0.7 | 0.1 |
| Externalizing scales | | | | | | |
| Hyperactive | 2.8 | 2.8 | 3.8 | 3.8 | 3.4 | 0.2 |
| Sex problems | 1.0 | 1.1 | 1.1 | 1.2 | 1.1 | 0.1 |
| Delinquent | 0.4 | 0.9 | 0.5 | 1.0 | 0.5 | 0.0 |
| Aggressive* | 7.2 | 6.0 | 8.5 | 7.3 | 8.3 | 0.3 |
| Cruel | 0.5 | 1.0 | 0.8 | 1.5 | 0.8 | 0.1 |
| Internalizing | 7.7 | 6.3 | 9.8 | 8.5 | 8.8 | 0.4 |
| Externalizing* | 10.7 | 8.6 | 13.1 | 11.0 | 12.6 | 0.5 |
| Total** | 19.9 | 14.2 | 24.9 | 18.8 | 23.6 | 0.9 |

T-test comparing CBCL sample means with twin sample unweighted means: * p < 0.05 ** p < 0.01

Table 5 - Mean scale scores for girls aged 12-16

| | CBCL | | Twin sample (Female raters only) | | | |
|-----------------------------|--------------------|-----------|----------------------------------|-----------|------------------|-----------|
| | Nonclinical sample | | Weighted means | | Unweighted means | |
| | Raw score | SD of raw | Raw score | SD of raw | Raw score | SE of raw |
| | N = 250 | | N = 378 | | N = 378 | |
| Internalizing scales | | | | | | |
| Anxious obsessive** | 3.6 | 3.7 | 4.8 | 4.6 | 4.7 | 0.2 |
| Somatic** | 0.6 | 1.1 | 1.5 | 2.0 | 1.5 | 0.1 |
| Schizoid | 0.9 | 1.2 | 0.9 | 1.2 | 0.9 | 0.1 |
| Depressed withdrawal** | 3.0 | 2.8 | 3.9 | 3.7 | 4.1 | 0.2 |
| Immature hyperactive* | 2.3 | 2.5 | 3.0 | 3.2 | 2.9 | 0.2 |
| Externalizing scales | | | | | | |
| Delinquent** | 2.4 | 3.3 | 3.4 | 3.7 | 3.3 | 0.2 |
| Aggressive** | 5.1 | 5.2 | 6.9 | 6.6 | 6.9 | 0.4 |
| Cruel** | 0.5 | 1.1 | 1.1 | 2.3 | 1.2 | 0.1 |
| Internalizing** | 7.0 | 6.5 | 10.0 | 8.8 | 10.0 | 0.5 |
| Externalizing** | 7.3 | 7.6 | 10.2 | 9.7 | 10.2 | 0.5 |
| Total** | 16.6 | 14.1 | 22.8 | 18.6 | 22.8 | 1.0 |

T-test comparing CBCL sample means with twin sample unweighted means: * p<0.05 ** p<0.01

DISCUSSION

The results of this study address issues surrounding generalization from twin data to the population. Data from the current analysis do not confirm the equivalence of twin samples and the general population. Though several of the internalizing subscales were not significantly different between the samples, the overall scores on externalizing and total behavior problem scales indicate significant differences between the twin sample and the non-twin sample. In addressing the more specific hypothesis regarding minimal brain damage in twins, the elevation in twins' total problem scores across age and gender groups provides tentative support for a relationship between perinatal insult and increased childhood behavior problems.

Clearly, there are other characteristics of the two samples, (eg, regional differences, method of data collection, subject selection) which require circumspection in interpreting these data. Taken at face value and in the absence of more clearcut comparison data, however, our data suggest some intriguing patterns of direct relevance to hypotheses about the etiology of some types of childhood psychopathology.

In summary, using a well standardized checklist assessment of childhood behavior problems in a population based twin sample, we have found that the mean scores were not equivalent for twins compared to a non-twin normative sample excluding children who had recently received mental health services. On the contrary, our evidence suggests that twins may show small but consistent elevations in behavior problem scales. These elevations were statistically significant for older children on both internalizing and externalizing behaviors; for younger children the elevations were significant for externalizing scales only.

Acknowledgements: The work reported in this paper was supported in part by grants MH45268 and AA08672, and the Carman Trust for Scientific Research.

REFERENCES

1. Achenbach TM, Edelbrock C (1983): *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington, VT: University of Vermont Department of Psychology.
2. Barkley RA (1990): *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment*. New York: Guilford.
3. Bulmer MG (1970): *The Biology of Twinning in Man*. Oxford: Clarendon.
4. Graham P, Stevenson J (1985): A twin study of genetic influences on behavioral deviance. *Am Acad Child Psychiatry*, 24: 33-41.
5. Hewitt JK (1990): Genetics as a framework for the study of behavioral development. In ME Hahn, JK Hewitt, ND Henderson, RH Benno (eds), *Developmental behavior genetics: Neural, biometrical and evolutionary approaches*. New York: Oxford University Press, pp. 298-303.
6. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ (in press): A population based twin study of major depression in women: The impact of varying definitions of illness. *Arch Gen Psychiatry*.
7. MacGillivray I, Campbell DM, Thompson B (1988): *Twinning and Twins*. Chichester: Wiley & Sons.

8. MacGillivray I, Nylander PPS, Corney G (1975): Human Multiple Reproduction. London: W.B. Saunders.
9. Pasamanick B, Knobloch H (1966): Retrospective studies on the epidemiology of reproductive casualty: old and new. *Merril-Palmer Behav Dev*, 12: 7-26.
10. Rutter M (1982): Syndromes attributed to "Minimal Brain Dysfunction" in childhood. *Am J Psychiatry*, 139: 21-33.
11. Silberg JL, Erickson MT, Meyer JM, Eaves LJ, Rutter ML, Hewitt JK (submitted): The contribution of genetic and environmental factors to maternal ratings of behavioral and emotional problems in adolescents. *J Consult Clin Psychology*.
12. Stevenson J, Graham P (1988): Behavioral deviance in 13-year-old twins: An item analysis. *J Am Acad Child Adolescent Psychiatry*, 27: 791-797.
13. Strauss AA, Kephart NC (1940): Behavior differences in mentally retarded children measured by a new behavior rating scale. *Am J Psychiatry*, 16: 46-53.

Correspondence: Dr. Jeanine S. Gau, Department of Psychology, Virginia Commonwealth University, Richmond, VA 23284-2018, USA.