Pergamon

J. Child Psychol. Psychiat. Vol. 36, No. 3, pp. 449–458, 1995
Elsevier Science Ltd

0021-9630/95 $9.50 + 0.00

0021–9630(94)00138–3

A Twin–Singleton Comparison of Problem Behaviour in 2–3-Year-Olds

Edwin J. C. G. van den Oord,*† Hans M. Koot,*
Dorret I. Boomsma,† Frank C. Verhulst* and J. F. Orlebeke†

Abstract—Twin–singleton differences in problem behaviours in 2–3-year-olds were studied. Maternal ratings of children's problem behaviours were obtained with the CBCL/2–3. The twin sample consisted of 1363 twin pairs (456 MZ, 907 DZ), the sample of singletons consisted of 420 children from the general population. Results indicated that the general level of problem behaviours in twins was broadly comparable to that in singletons. Four of the seven syndromes showed lower scores for twins. These differences, however, were small and mainly caused by lower scores for DZ twins in comparison to MZ twins and singletons. Part of the difference could be attributed to the higher maternal age in the twin groups. Higher means for boys were found for the total problem score, and the Aggressive and Overactive syndromes.

Keywords: Child psychopathology, twins, Child Behavior Checklist
Abbreviations: CBCL: Child Behavior Checklist, MZ: monozygotic, DZ: dizygotic

Introduction

Twin studies are frequently used to study genetic influences on problem behaviours in children. However, the generalization of findings from twin studies to the general population may be limited by differences between twin and nontwin samples (Bryan, 1993).

There are 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, a lower birth weight, and a shorter length of gestation (Bulmer, 1970, p. 46; Little & Bryan, 1988; MacGillivray & Campbell, 1988).

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 nontwins, because parents of twins

Accepted manuscript received 5 *October* 1994

*Department of Child and Adolescent Psychiatry, Sophia Children's Hospital, Erasmus University, Rotterdam, The Netherlands.
†Department of Psychonomics, Free University, Amsterdam, The Netherlands.
Requests for reprints to: Edwin van den Oord, Czaar Peterstraat 35[s], 1018 NX, Amsterdam, The Netherlands.

have to divide their resources between two children of a comparable developmental level. Thorpe, Golding, MacGillivray and Greenwood (1991), for example, showed that mothers of twins are more likely to experience depression which in turn could affect the child's development. Interactions between the twins themselves might also be a source of twin–singleton differences.

There are a few studies that have compared problem behaviours in twins and singletons in middle childhood and adolescence. Most studies showed only small differences, and suggested that the level of problem behaviour is broadly comparable in both groups (Hay & O'Brien, 1984, 1987; Simonoff, 1992;). In contrast, a recent study by Gau, Silberg, Erickson and Hewitt (1992) showed small but consistently higher levels of problem behaviours in twins compared with singletons.

For children of preschool age, even less is known about the level of problem behaviour in twin versus nontwin samples. Kim, Dales, Connor, Walters and Witherspoon (1969), employing a sample of 13 pairs of MZ twins and 22 singletons, reported lower levels of aggressive behaviours in twins than in singletons. Lytton, Conway and Sauvé (1977), and Lytton (1980, p. 157) found 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 behaviours that were addressed in these two studies, make it difficult to draw firm conclusions concerning twin/singleton differences in problem behaviours in preschool children.

The aim of the present paper was to study twin–singleton differences in problem behaviours in 2–3-year old children. Maternal ratings of problem behaviours in twins (1363 pairs) and singletons (420 children) were obtained with the CBCL for ages 2–3 years (Achenbach, 1992). Mean problem scores and standard deviations were compared for groups of MZ, DZ, and singletons.

Materials and Methods

Measure

The CBCL/2–3 (Achenbach, 1992) is an assessment instrument to obtain parental ratings of problem behaviours in 2–3-year-olds. The CBCL/2–3 was modeled after 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 (1993). The Dutch syndromes differ somewhat from those reported for American samples (Achenbach, 1992), and are labeled Oppositional (items 8, 13, 16, 29, 30, 33, 36, 44, 66, 69, 81, 82, 83, 85, 88, 96, 97), Withdrawn/Depressed (items 2, 23, 26, 43, 67, 70, 71, 77, 80, 90), Aggressive Behaviour (items 14, 17, 18, 20, 35, 40, 42, 53, 91), Anxious (items 3, 4, 10, 21, 37, 68, 73, 87, 92), Overactive (items 5, 6, 11, 59, 62), and Sleep Problems (items 22, 38, 48, 64, 74, 84, 94). A seventh syndrome labeled Somatic Problems was not included in the present study because it could not be reliably assessed, and frequencies of problems comprising this syndrome were very low in both the twin and community sample.

In addition to these syndromes, the total problem score was studied. The total problem score is the sum of all 99 items, and includes items that do not appear in one of the syndrome scales. It can be viewed as an overall index of the number and severity of reported problems.

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

Netherlands 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. Nonresponders were sent reminders and, when no response was obtained, contacted by phone. Completed questionnaires were returned by 1377 parents (77%), which is about (45% × 77%) 35% of all Dutch twins in the target age.

For 408 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 560 same sex twin pairs information about zygosity was obtained from a questionnaire completed by parents when in almost all cases, the twins were about 2 years old. Twenty-six families indicated that they were not certain about the zygosity of their twin. These parents were contacted by phone. Fourteen twin pairs were discarded because their parents were still uncertain. This procedure left a sample of 242 MZ female, 214 MZ male, 235 DZ female, 263 DZ male, and 409 opposite-sex pairs.

To establish the reliability of the questionnaire used to determine the twin's zygosity, blood test results were compared with the zygosity information from the questionnaire. For 356 same-sex twin pairs both blood test and questionnaire results were available. The agreement between the two measures was 82%. It could very well be that parents who were uncertain about their twin's zygosity were more likely to consent to a blood test. This percentage should probably therefore be viewed as the lower bound of the reliability of the questionnaire. More MZ twins appeared to be misclassified as DZ (52 pairs) than DZ twins as MZ (14 pairs).

Community sample. Subjects in the community sample were 420 children (215 boys, 205 girls) from a target sample of 469 children (90% response rate). They were randomly drawn from the inoculation register, which includes 95% of all children aged 2–3 years, 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 female interviewers in order to obtain the highest possible response rate. The interviewer read the first CBCL/2–3 problem items regarding the target child aloud, then the parents proceeded themselves and the interviewer scored the responses.

Data analyses

First, the twin and community sample were compared on a number of background characteristics using chi-squared probability tests for categorial variables and ANOVAs for continuously distributed background variables. Then 2-way ANOVAs with twin–singleton/girl–boy effects were used to test for group differences in syndrome scores and interaction effects. Finally, ANCOVAs were used to study whether differences in syndrome scores could be explained by differences in background characteristics for which significant differences between the community and twin sample were found.

Where significant differences between twins and singletons in background characteristics and syndrome scores were found, we examined whether MZ twins differed from DZ twins, MZ twins differed from singletons, or whether DZ twins differed from singletons.

Data obtained for pairs of twins are not independent. For instance, background characteristics such as age and parental occupation are identical, and syndrome scores will be associated because twins share genes and environments. To remove this dependence from the data and perform accurate significance tests, one twin was randomly selected from each pair. For all the tests, $p < .05$ was used (This method gave essentially identical results to a more complex approach using LISREL, which enables one to retain all the data in the analysis; see Appendix 1).

Maternal ratings, which were available for 99.1% of the twins and 98.3% of the children in the community sample, were used to study group differences on syndrome scores. To obtain an impression of the magnitude of the differences that were found, Cohen's (1988, p. 20) effect size was computed by dividing the absolute difference between the group means 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.

Results

Demographic characteristics

Demographic characteristics of the twin and community samples are shown in Table 1.

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

	Twin Sample		Community sample	
	Monozygotic	Dizygotic		$p<$
	pairs = 456	pairs = 907	N = 420	
Age of child (in months)	42.0 (3.91)	42.2 (3.99)	36.4 (7.06)	.000
Age of the parents (in years)				
Mother	32.7 (4.09)	33.3 (3.75)	31.5 (4.41)	.000
Father	35.3 (5.08)	36.0 (4.33)	34.1 (5.04)	.000
Maternal parity	1.87 (.93)	1.91 (.98)	1.82 (.90)	.215
Paid labour				
Mother	29.5%	28.8%	27.1%	.681
Father	97.8%	96.9%	92.5%	.000
Level of parental occupation				
Mother	3.58 (1.34)	3.64 (1.38)	3.53 (1.41)	.826
Father	3.51 (1.39)	3.54 (1.43)	3.68 (1.44)	.319

Standard deviations are in parentheses. The level of parental occupation (ITS, Van Westerlaak, Kropman & Collaris, 1975) is the mean of a six-step scale (where 6 = highest level). In the twin sample parity information was available for 1091 of the 1363 families.

There were no twin/singleton differences in maternal parity, the occupational level of mothers and fathers, and the employment rate of the mothers. For the age of the child, the age of the mothers and fathers, and whether or not the fathers were employed there were significant differences between twins and singletons.

Pairwise comparisons showed for most background characteristics that both twin groups differed from singletons, but indicated no differences between MZ and DZ twins. Twins were 5–6 months older than singletons, parents of twins were 1.35 years older than parents of singletons, and the paternal employment rate was 4.9% higher in the twin sample. Age of the mother was the only background characteristic which showed a significant difference ($p < .014$) between MZ and DZ twins. Mothers of DZ twins were somewhat older than mothers of MZ twins. Although a similar trend was present for the paternal age, it just fell short of significance ($p < .082$).

ANOVAs

In preparation for the ANOVAs we examined whether levels of problem behaviour were comparable for first- and second-borns. For instance, the higher risk for birth injury in second-born twins or their lower birth weights (Bulmer, 1970, pp. 62–64; Campbell & MacGillivray, 1988; Orlebeke, Van Baal, Boomsma & Neeleman, 1993), could result in higher syndrome scores in this group. To check for these kind of birth order effects, *t*-tests for paired samples were performed using the data of both twins from the 954 same-sex pairs. The total problem score showed somewhat higher syndrome scores for the first born twin ($p < .045$). However, results for all other syndromes were nonsignificant, indicating no differences among first- and second-born twins.

Table 2 shows the means and standard deviations for the maternal ratings of twins and singletons, and presents the results from the 2-way ANOVAs. Except for the Withdrawn/Depressed syndrome, differences between mean syndrome scores of twins and singletons were significant. For the total problem score, Oppositional, Overactive, and Sleep Problems these differences consisted of lower scores for twins, for the Anxious syndrome it consisted of higher scores for twins than for singletons. According to Cohen's criteria (1988, p. 20), however, these differences were small.

Sex differences were found for the total problem score, Aggressive Behavior, and the Overactive syndrome. 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.

There were no significant interactions between twin–singleton/girl–boy effects for any of the syndromes. Significance levels ranged from $p < .262$ for the Anxious syndrome to $p < .714$ for Aggressive Behavior.

Pairwise comparisons showed for the total problem score, Oppositional, Aggressive, Overactive, and Sleep Problems significantly lower scores for DZ twins compared with MZ twins and singletons, but no significant differences among MZ twins and singletons. For the Anxious syndrome MZ twins obtained significantly higher scores than singletons, for the Withdrawn/Depressed syndrome there were no significant differences at all.

ANCOVAs

Table 1 showed significant differences between twins and singletons for age of the child, age of both parents, and paternal employment rate. In addition, mothers of DZ twins were older than mothers of MZ twins. These background characteristics were included in the ANCOVAs, to examine whether differences in syndrome scores could be attributed to differences in these background characteristics.

Associations between background characteristics and problem scores were small. The most important covariate was the age of the mother, which had a significant effect on all syndromes. For the total problem score, Oppositional, Aggressive, and the Overactive syndrome the standardized regression coefficients were close to –.14. Regression coefficients for the other syndromes were all smaller than –.10. The age of the child had a significant effect on the Withdrawn/Depressed and Overactive syndromes (standardized regression coefficients were, respectively .07

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

	MZ-twins		DZ-twins		Normative sample		
	Girls	Boys	Girls	Boys	Girls	Boys	ANOVA
	$N = 484$	$N = 428$	$N = 879$	$N = 935$	$N = 199$	$N = 214$	
Total problem score							
Mean	31.5	33.4	26.7	30.9	32.3	34.4	$p < .000$, twins < sing., d = .18
SD	19.9	18.6	17.0	18.6	16.6	17.0	$p < .001$, girls < boys, d = .17
Oppositional							
Mean	10.3	10.5	9.14	9.75	10.8	10.6	$p < .001$, twins < sing., d = .14
SD	7.04	6.44	6.34	6.68	5.97	6.03	$p < .518$, girls = boys
Withdrawn/Depressed							
Mean	1.11	1.19	1.06	1.18	1.02	1.26	$p < .936$, twins = sing.
SD	1.42	1.46	1.59	1.69	1.54	2.09	$p < .185$, girls = boys
Aggressive							
Mean	2.98	4.15	2.37	3.76	2.61	3.71	$p < .007$, DZ < sing. < MZ
SD	2.63	3.06	2.36	2.97	2.04	2.94	$p < .000$, girls < boys, d = .44
Anxious							
Mean	3.93	3.50	3.28	3.52	3.27	3.26	$p = < .044$, twins > sing., d = .08
SD	3.60	2.87	3.03	3.19	2.80	3.03	$p < .518$, girls = boys
Overactive							
Mean	2.62	3.09	2.35	2.76	3.05	3.24	$p < .000$, twins < sing., d = .22
SD	2.13	2.28	2.19	2.25	2.25	2.60	$p < .001$, girls < boys, d = .17
Sleep problems							
Mean	2.15	1.89	1.62	1.67	2.21	2.07	$p < .001$, twins < sing., d = .17
SD	2.40	2.06	2.10	2.05	2.53	2.57	$p < .827$, girls = boys

Sing. is singletons, *SD* is standard deviation, d is Cohen's (1988, p. 20) effect size. Mean and *SD* were computed using the ratings of all twins, for the ANOVAs one twin was randomly selected from each pair.

and –.05). The paternal employment rate had a significant effect on the total problem score, Oppositional, and Overactive syndrome, but again the regression coefficients of about –.05 indicated only very small effects. Age of the father did not have a significant effect on any of the scales.

For the Oppositional ($p < .147$) and Anxious syndromes ($p < .148$) the ANCOVAs indicated nonsignificant differences between twins and singletons. The inclusion of covariates also tended to result in nonsignificant differences for Overactive ($p < .049$), Sleep Problems ($p < .038$), and the total problem score ($p < .013$). These findings suggested that part of the differences that were found between twins and singletons may be attributed to differences in background characteristics, most importantly the age of the mother.

For the twin sample additional background information was available consisting of whether or not the child had physical problems, had been admitted to a hospital, was anaesthetized, or if a doctor had been consulted. Chi-square tests indicated no significant differences between MZ and DZ twins on these characteristics, and it is therefore unlikely that the higher levels of problem behaviours found for some scales for MZ twins can be attributed to any of these characteristics.

Discussion

Twin–singleton differences in problem behaviours in 2–3-year-olds were studied. Results indicated that the level of problem behaviours in twins was broadly comparable to that in singletons. Four of the seven syndromes showed lower scores for twins. However, these differences were small and were mainly caused by lower scores for DZ twins in comparison to MZ twins and singletons. Part of the difference could be attributed to the higher maternal age in the twin groups. Higher means for boys were found for the total problem score, and the Aggressive and Overactive syndromes.

Results from the present study were in agreement with most studies that have compared problem behaviours in twins and singletons in middle childhood and adolescence. However, they do not agree with the CBCL study by Gau *et al.* (1992), which showed small but consistently higher levels of problem behaviours in twins than in singletons. Gau *et al.* (1992) compared maternal ratings of problem behaviours in 1824 adolescent twins with an American normative sample, which consisted of a community sample with the exclusion of children who recently had received mental health services (Achenbach & Edelbrock, 1983). A number of reasons such as the age difference, the use of an American versus a Dutch sample, the different percentage of families that returned the questionnaire (44% versus 77% 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 response rate was higher in the community sample than in the twin sample (90% versus 77%). Part of this difference can be explained by the larger numbers of untraceables in the twin sample. In addition, the parents of singletons were visited by an interviewer, but questionnaires were mailed to the parents of twins. It could be that parents of problem children were more reluctant to return the questionnaire, explaining the lower means for four of the seven scales in the twin sample. We therefore compared the ratings of the 1075 parents who returned the questionnaire immediately, with the ratings of the 302 parents who completed the questionnaire after they were sent reminders and/or contacted by phone. Only for the Anxious syndrome there was a significant difference ($p < .01$), but means in the latter group were lower instead of higher. These findings did not suggest higher levels of problem behaviours in children of parents who were more reluctant to return the questionnaires.

Results from the present study suggest that in spite of the fact that twins have higher rates of biological and psycho-social hazards (Thorpe *et al.*, 1991) that may be important in individual cases they seem to have a limited impact on twins as a group. This is probably both because most twins are physically healthy and grow up in normal circumstances and because the psychological risks associated with these factors are quite small (Rutter & Redshaw, 1991).

This conclusion is also consistent with two other findings in the present study. Firstly, no interaction effects were found in the two-way ANOVAs. For children born with birth weights less than 1500 grams, for instance, a number of studies have reported an increase of behaviour problems in boys, but not in girls (Breslau, Klein & Alen, 1988; Ross, Lipper & Auld, 1990). This finding would suggest a 'twin effect' especially in boys. However, such an interaction effect would remain undetected in a twin sample that comprises only a small proportion of low birth weight children. For example, Hay and O'Brien (1984), employing a sample of over 200 twins, found that 2.4% of the twins were born weighing less than 1500 grams.

Secondly, no evidence was found of birth order effects. The improvements in the management of twins gestation, delivery and care in the neonatal period not only have reduced the biological risks for twins in general but also decreased differences between first- and second-born twins (Campbell & MacGillivray, 1988; Depp, Keith & Sciarra, 1988). These differences, that probably have only a minor impact on the psychological functioning of the twins, would therefore have little influence on a comparison between mainly healthy twins.

Somewhat lower means for DZ twins compared to MZ twins were found for five of the seven scales. The differences, however, were small according to Cohen's criteria (1988, p. 20) and could in part be explained by the higher age of the mothers of DZ twins. It is known that mothers of DZ twins are older than mothers of MZ twins (Bulmer, 1970, p. 80), and its effect on problem behaviours studied in the present paper therefore seems to represent a meaningful effect. It should, however, not be interpreted as a parity effect, since no difference was found in the parity of MZ and DZ mothers.

The present study showed few differences in the distributions of problem scores in 2–3-year-old twins and singletons. These findings therefore tend to provide support for the generalizability of findings from twin studies, at least for twins of preschool age, to the general population.

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

References

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

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

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

Bulmer, M. G. (1970). *The biology of twinning in man.* Oxford: Clarendon Press.

Breslau, N., Klein, N. & Alen, L. (1988). Very low birthweight: behavioral sequela at nine years of age. *Journal of the American Academy of Child and Adolescent Psychiatry,* **27,** 605–611.

Bryan, E. M. (1993). Prenatal and perinatal influences on twin children: implications for behavioral studies. In T. J. Bouchard & P. Propping (Eds.), *Twins as a tool of behavioral genetics* (pp. 209–217). New York: Wiley.

Campbell, D. M. & MacGillivray, I. (1988). Outcome of twin pregnancies. In I. MacGillivray, D. M. Campbell & B. Thompson (Eds.), *Twinning and twins* (pp. 179–202). Chichester: Wiley.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Depp, R., Keith, L. G. & Sciarra, J. J. (1988). The Northwestern University Twin Study VII: the mode of delivery in twin pregnancy North American considerations. *Acta Geneticae Medicae et Gemellologiae,* **37,** 11–18.

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.

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.

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

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.

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

Koot, H. M. & Verhulst, F. C. (1991). Prevalence of problem behavior in Dutch children aged 2–3. *Acta Psychiatrica Scandinavica,* **83** (suppl. 367), 1–37.

Little, J. & Bryan, E. M. (1988). Congenital anomalies. In I. MacGillivray, D. M. Campbell & B. Thompson (Eds), *Twinning and twins* (pp. 207–240). Chichester: Wiley.

Lytton, H. (1980). *Parent–child interaction: the socialization process observed in twin–singleton families.* New York: Plenum Press.

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

MacGillivray, I. & Campbell, D. M. (1988). Factors affecting twinning. In I. MacGillivray, D. M. Campbell & B. Thompson (Eds.), *Twinning and twins* (pp. 111–142). Chichester: Wiley.

Orlebeke, J. F., Van Baal, G. C. M., Boomsma, D. I. & Neeleman (1993). Birth weight in opposite sex twins as compared to same sex dizygotic twins. *European Journal of Obstetrics and Gynecology and Reproductive Biology,* **50,** 95–98.

Ross, G., Lipper, E. G. & Auld, P. A. M. (1990). Social competence and behavior problems in premature children at school age. *Pediatrics,* **86,** 391–397.

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.

Simonoff, E. (1992). A comparison of twins and singletons with child psychiatric disorders. *Journal of Child Psychology and Psychiatry,* **33,** 1319–1332.

Thorpe, K., Golding, J., MacGillivray, I. & Greenwood, R. (1991). Comparison of prevalence of depression in mothers of twins and mothers of singletons. *British Medical Journal*, **320**, 875–878.
Van den Oord, E. J. C. G. (1993). *A genetic study of problem behaviors in children.* Unpublished doctoral dissertation. Eramus University Rotterdam.
Van Westerlaak, J. M., Kropman, J. A. & Collaris, J. W. M. (1975). *Beroepenklapper [Index of occupations].* Nijmegen: Institut voor Toegepaste Sociologie.

Appendix 1

Tests were also performed using a multi-sample analysis (Jöreskog & Sörbom, 1989, pp. 227–244) in LISREL, which enables one to retain all the data in the analysis. In a multi-sample analysis the fit of models assuming that means are not equal across groups can be compared with the fit of models that constrain means to be equal. A significant worse fit of the latter model indicates that means differ across groups.

The test were performed in LISREL by specifying in the groups of twins two variables, one for each child. However, because LISREL requires that every group has the same number of variables and there is only one 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 also included for this group (analogous to the way missing data can be handled in LISREL, Jöreskog & Sörbom, 1989, p. 259). For the seven groups in the analysis (MZ girls, MZ boys, DZ girls, DZ boys, DZ opposite sex, singletons boys, singletons girls) there were 29 real statistics and six statistics associated with the dummy variables. To perform accurate significance tests the degrees of freedom were adjusted for these dummy statistics by putting $df = -6$ on the OU line of the last group.

These analyses with LISREL yielded essentially the same result as the ANOVA's that were used in the paper (Van den Oord, 1993). Thus, the pattern of significant and nonsignificant results for the tests of differences in mean levels of problem behaviours between twins and singletons was the same for both techniques.