A Twin-Singleton Comparison of Developmental Trajectories of Externalizing and Internalizing Problems in 6- to 12-Year-Old Children

Sylvana C. C. Robbers,[1,2] Meike Bartels,[2] Floor V. A. van Oort,[1] C. E. M. (Toos) van Beijsterveldt,[2] Jan van der Ende,[1] Frank C. Verhulst,[1] Dorret I. Boomsma,[2] and Anja C. Huizink[1,3]

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

[2] *Department of Biological Psychology, VU University Amsterdam, The Netherlands*

[3] *Department of Education, Faculty of Behavioral and Social Sciences, University of Amsterdam, The Netherlands*

Research on twin-singleton differences in externalizing and internalizing problems in childhood is largely cross-sectional and yields contrasting results. The goal of this study was to compare developmental trajectories of externalizing and internalizing problems in 6- to 12-year-old twins and singletons. Child Behavior Checklist (CBCL) maternal reports of externalizing and internalizing problems were obtained for a sample of 9651 twins from the Netherlands Twin Register and for a representative general population sample of 1351 singletons. Latent growth modeling was applied to estimate growth curves for twins and singletons. Twin-singleton differences in the intercepts and slopes of the growth curves were examined. The developmental trajectories of externalizing problems showed a linear decrease over time, and were not significantly different for twins and singletons. Internalizing problems seem to develop similarly for twins and singletons up to age 9. After this age twins' internalizing symptoms start to decrease in comparison to those of singletons, resulting in less internalizing problems than singletons by the age of 12 years. Our findings confirm the generalizability of twin studies to singleton populations with regard to externalizing problems in middle and late childhood. The generalizability of studies on internalizing problems in early adolescence in twin samples should be addressed with care. Twinship may be a protective factor in the development of internalizing problems during early adolescence.

Keywords: child externalizing and internalizing problems, twins, singletons

Twins are frequently used to study causes of individual differences in problem behaviors. An important assumption of twin studies is that results from twin samples can be generalized to the general population. However, the comparability of twins to singletons is still being questioned for various reasons. For instance, increased pre- and perinatal risks among twins could result in a higher prevalence of behavioral and emotional problems in twins than in singletons (Rutter & Redshaw, 1991). Twins are born on average three weeks earlier than singletons, and their birth weight is on average lower (Rutter & Redshaw, 1991). Several authors have reported increased levels of problems such as ADHD (Lahti et al., 2006) and depression (Raikkonen et al., 2008) in preterm children, children with low birthweight and children with small body size at birth. However, low birth weight, preterm birth, or small size at birth are unlikely to have the same significance in twins as in singletons, since the etiology of these risk factors appears to be different in the two groups (Phillips et al., 2001). In twins, small size at birth results from 'crowding', rather than from an unfavorable intrauterine environment (Gruenwald, 1970). From about 24 weeks gestational age onwards placental weights of twins are smaller than those of singletons. Children in multiple gestations become growth retarded, preceded by and very likely due to poor early placental development (Bleker et al., 1988). After birth, twins grow fairly well compared to singletons, and twins are almost as tall as singletons by the age of 5 years (Estourgie-van Burk et al., 2006). It has also been hypothesized that twins may show fewer problems than singletons because the twinship offers a favorable social environment, via interactions with, and social support of the co-twin (Pulkkinen et al., 2003), although there is little evidence supporting this hypothesis as yet.

Despite the importance and ongoing debate about the generalizability of twin studies, the number of

Received 02 October, 2009; accepted 16 October, 2009.

Address for correspondence: Dr. A.C. Huizink, Erasmus MC-Sophia Children's Hospital, Department of Child and Adolescent Psychiatry, PO box 2060, 3000 CB Rotterdam, The Netherlands. Email: a.c.huizink@erasmusmc.nl.

studies that investigated twin-singleton differences in problem behaviors is surprisingly limited. Researchers have so far focused on comparing mean levels of externalizing and internalizing problems between twins and singletons in childhood, and their findings are mixed. For instance, a study with 8- and 9-year-old Finnish twins and singletons showed significantly lower rates of externalizing problems in twins (Moilanen et al., 1999). Similarly, slightly less externalizing problems were found in 2- and 3-year-old Dutch twins compared to same-aged singletons (Van den Oord et al., 1995). In contrast, higher levels of externalizing problems in twins than in singletons were reported in an American sample of 6- to 16-year-old children (Gau et al., 1992). Also, 4- to 12-year-old twins showed more ADHD symptoms than singletons (Levy et al., 1996) and twins showed more conduct disorders than singletons (Simonoff, 1992). Studies among 11- and 12-year-old Finnish twins and singletons (Pulkkinen et al., 2003), and 5- to 15-year-old Norwegian twins and singletons (Gjone & Novik, 1995) showed no evidence for twin-singleton differences in externalizing problems. The level of internalizing problems in Norwegian children aged 5 to 15 years has been found to be lower in twins than in singletons, especially around the age of 12 and most pronounced for boys (Gjone & Novik, 1995). Other studies found no evidence for twin-singleton differences in internalizing problems until the age of 12 (Gau et al., 1992; Pulkkinen et al., 2003).

The results of these studies may be conflicting for various reasons, such as the use of different measures, sample sizes, and age groups. A major constraint has been that all studies conducted so far are cross-sectional. Therefore, further investigation regarding whether or not results from twin samples can be generalized to singletons is needed. Information is lacking on twin-singleton differences in the longitudinal development of problem behaviors in childhood. Therefore, the aim of the present study is to compare growth curves of the development of externalizing and internalizing problems, using 6- to 12-year-old twins and singletons. In addition to cross-sectional mean differences, growth curves may yield twin-singleton differences in the overall development over time that would otherwise remain undetected. Previous studies on the normative development of internalizing and externalizing behaviors in childhood show declining trajectories of externalizing problems for both sexes (Bongers et al., 2003; Leve et al., 2005; Miner & Clarke-Stewart, 2008; Stanger et al., 1997), with boys having higher mean levels of externalizing problems than girls. Regarding the development of internalizing problems in childhood, the results are somewhat mixed. Previous studies reported increasing trajectories for girls, but not for boys (Leve et al., 2005), stable trajectories for both boys and girls (Keiley et al., 2000), and curvilinear increasing trajectories for both boys and girls (Bongers et al., 2003), with boys and girls having similar mean levels of internalizing problems in childhood. We expect to observe a similar decrease of externalizing problems and an increase of internalizing problems over time for twins and singletons.

Materials and Methods

Subjects

Singleton sample. The singleton data used in this study were derived from an ongoing Dutch longitudinal study of behavioral and emotional problems that started in 1983. The original sample of 2600 children aged 4 to 16 was randomly drawn from municipal registers that list all residents in the Dutch province of Zuid-Holland, and represents a general population sample (Verhulst et al., 1985). After complete description of the study to the subjects, written informed consent was obtained. For details about the initial data collection, see Verhulst et al. (1985). The present study uses data from the first five waves. After the first measurement in 1983, the respondents were approached again in 1985, 1987, 1989 and 1991. At study onset, 2076 out of 2600 invited participants (80%) responded. Response rates ranged from 80% to 85% at each measurement. For 1964 children, within an age range of 4 to 18, mother ratings were available on at least one of the five waves of data collection. These children were all born between 1971 and 1979. We kept all subjects in the sample who were between 6 and 12 years of age at any time point, even if mother ratings were available from only one measurement. This resulted in a final singleton sample of 1351 children (664 boys and 687 girls). For 39% of this sample we had data from one measurement, for 26% from two measurements, for 23% from three measurements, and for 12% from four measurements. Due to the design of this study, the number of measurements of each child was not only determined by dropping out, but it was restricted by birth cohort. For example, children born in 1971 only had data at age 12, while children born in 1979 had data at multiple points in time. Attrition analyses on the initial 2076 respondents revealed a significant effect of dropout on socio-economic status (SES), with dropouts having lower SES. There was no effect of dropout on the mean level of problems (Bongers et al., 2003).

Twin sample. The twin data came from a longitudinal twin study on health, growth, and the development of behavioral and emotional problems. All participating families are volunteer members of the Netherlands Twin Register (NTR), established by the Department of Biological Psychology at the VU University in Amsterdam, and represent a twin family sample that is representative for the Dutch general population (Bartels et al., 2007; Boomsma et al., 2006). The study protocol was approved by the ethics committee of the VUMC (VU University medical center), Amsterdam. Of all multiple births in the Netherlands, 40% to 50% is registered by the NTR since 1986 (Boomsma et al., 1992; Boomsma et al., 2002). For the present study, data from twins born between 1986 and 1998 were

analyzed. Parents of twins received surveys by mail, around the twins' 7th, 10th and 12th birthdays. The absolute response rate at each measurement was about 62%. Attrition analyses revealed that dropout was not related to problem behavior (Bartels et al., 2007). Although most twins were aged 7, 10, or 12, there were also a substantial number of children who were older or younger than these target ages at the time of the assessments, and by calculating the exact ages in years of the twins at the moment of completion of the questionnaires data from 6 to 12 years were available for analyses. Data were available for 19274 twins from 9651 families. Since data obtained from twin pairs are not independent, one twin was randomly selected from each pair. This resulted in a final sample of 9651 twins (4728 boys and 4923 girls). For 43% of this twin sample data were available from one measurement, for 28% from two measurements, and for 29% from three measurements. Children with longitudinal data were in general from older birth cohorts, since younger cohorts did not reach all target ages of survey collection yet (e.g., twins born in 1998 are not yet invited for age 12 participation).

Covariates

It has been documented that economic problems have an adverse influence on the behavioral and emotional development of children (McLoyd, 1998). To make sure that any observed twin-singleton difference could not be attributed to differences in SES between our samples, SES was included as a covariate. For the singleton sample, SES was scored on a six-step scale of parental occupation (Van Westerlaak et al., 1975) with 1 indicating the lowest SES and 6 indicating the highest SES. This scale was subsequently divided into three SES levels (1 to 3 = low SES; 4 and 5 = middle SES; 6 = high SES). For 5 singletons information on SES was missing. For about two-third of the twin sample, SES was obtained from a full description of the occupation of the parents, and SES was subsequently coded according to the Dutch Standard Classification of Occupations (CBS, 2001). For the remaining twins, SES was obtained by the EPG-classification scheme (Erikson et al., 1979), combined with information on parental education. For all twins, the level of occupation was classified into SES levels that were similar to the ones used in the singleton sample. For 73 twins SES status was unknown. For both singletons and twins we used the baseline SES status, assessed at the first measurement occasion. Because the development of problem behaviors is different for boys than for girls (Bongers et al., 2003), we also included sex as a covariate to account for gender differences.

Instrument

For both the singletons and the twins, mother ratings of externalizing and internalizing problems were collected using the Child Behavior Checklist (CBCL/4-18; Achenbach, 1991; Achenbach, 1992). The CBCL/4-18 was developed for parents to rate the behavioral and emotional problems of their children. It consists of 120 items that are scored on a 3-point scale based on the occurrence of the behavior during the preceding 6 months: 0 if the item was not true, 1 if the item was somewhat or sometimes true, and 2 if the item was very true or often true. The good reliability and validity of the CBCL/4-18 were confirmed for the Dutch translation of the measure: Cronbach's alpha was .86 for the Externalizing scale and .83 for the Internalizing scale (Verhulst et al., 1996). The Internalizing scale consists of the Anxious/Depressed, Somatic Complaints, and Withdrawn subscales, and consists of 31 items. The Externalizing scale consists of the Aggressive Behavior and Rule-Breaking Behavior subscales, and consists of 33 items.

Data Analysis

Means and standard deviations of the problem scores were computed per age, using SPSS15. The developmental trajectories of externalizing and internalizing problems were examined using latent growth curve modeling (LGM; McArdle & Erpstein, 1987) with Mplus Version 5 (Muthén & Muthén, 1998–2007). In LGM, random effects are used to capture individual differences in development. These random effects are conceptualized as continuous latent variables, the growth factors. The growth curves were determined by two or three latent growth factors: (a) the intercepts, which represent the initial status of the curve; (b) the linear slopes, which represent linear change over time; and (c) the quadratic slopes, which represent non-linear change. We included quadratic slopes because nonlinear development of problem behaviors has been reported by earlier studies (Bongers et al., 2003). The covariates were centered to their means. Maximum likelihood was used to estimate the growth models. We allowed for the analysis of respondents with missing data, because in LGM a person who participates only once or whose SES status is unknown, still contributes to the overall estimation of the model. We performed multi-group analyses (i.e., singletons and twins) using the grouping option in Mplus. All analyses were conducted for externalizing and internalizing problems separately.

Evaluation of the growth curve models was conducted as follows. First, we examined the model fit, which was evaluated with two goodness-of-fit indices (Hu & Bentler, 1999): the comparative fit index (CFI), with values of >.95 indicating a good fit, and the root mean square error of approximation (RMSEA), with values of <.06 indicating a good fit. Second, we checked if the latent growth factor means and variances were significant and if the covariates had significant influence on the growth factor means. Third, we performed likelihood-based chi-square difference tests in Mplus to test whether the growth factor means and variances were different between twins and singletons. A growth model in which the means or variances of either the intercepts or the slopes were constrained to be equal for twins and singletons,

was tested against the unconstrained growth model. The linear and the quadratic slopes were jointly constrained. We tested if twin-singleton differences were different for boys and girls. This was not the case; therefore, the data were further analyzed for boys and girls together, including SES and sex as covariates.

Results

A total of 447 singletons (33%) had low SES, 444 singletons (33%) had middle SES, and 455 singletons (34%) had high SES. A total of 2043 twins (21%) had low SES, 4245 twins (44%) had middle SES and 3290 twins (34%) had high SES. The twin sample had a lower proportion of low SES children than the singleton sample. Sex was equally distributed in both samples ($\chi^2(1) = .012$, $p = .91$), with 51% boys in each sample. Table 1 shows the observed means and standard deviations for both twins and singletons for externalizing and internalizing problems, separately for boys and girls. Significant differences in these mean scores between twins and singletons were only observed for boys' externalizing problems at age 7 ($p = .03$), for boys' internalizing problems at age 11 ($p = .03$) and age 12 ($p = .01$), and for girls' internalizing problems at age 12 ($p = .02$). Singletons had higher problem scores than twins at these ages.

The model fit statistics and model results for externalizing problems are reported in Table 2. The model fit the data well (CFI = .99 and RMSEA = .014). The influences of sex and SES on the intercept were significant ($p < .001$) and negative for both twins and singletons, implying that children with high SES showed less externalizing problems, and that boys had more externalizing problems than girls. The effect sizes for sex differences in the intercepts of externalizing problems were moderate in both samples (Cohen's d singletons/twins = .50/.38). No significant effects of sex and SES were found on the slope for both twins

Table 1

Means and Standard Deviations of Externalizing and Internalizing Problem Scores for Singletons (664 Boys and 687 Girls) and Twins (4728 Boys and 4923 Girls)

			Singletons			Twins		
	Sex	Age	*N*	Mean	Std.	*N*	Mean	Std.
Externalizing	Boys	6	111	9.52	8.50	274	8.80	7.18
		7	121	10.33*	7.57	3644	8.76*	7.23
		8	177	8.42	7.19	332	8.87	7.25
		9	183	8.80	7.59	1205	8.24	7.68
		10	231	8.17	7.13	1511	7.96	7.15
		11	232	8.06	7.18	809	6.92	7.11
		12	306	7.16	7.01	975	7.27	7.13
	Girls	6	113	7.08	5.91	331	6.99	6.28
		7	136	5.98	5.63	3709	6.44	5.98
		8	179	7.01	6.20	370	6.76	6.32
		9	201	5.65	5.94	1378	5.75	5.80
		10	238	5.61	5.32	1561	5.67	5.99
		11	273	5.27	5.78	880	4.83	5.07
		12	296	5.44	5.49	1017	4.91	5.59
Internalizing	Boys	6	111	3.52	3.59	268	4.15	3.98
		7	121	4.83	4.80	3582	4.43	4.52
		8	177	4.61	4.49	323	4.39	4.36
		9	183	5.04	5.06	1192	4.77	4.79
		10	231	5.35	5.13	1487	4.68	5.02
		11	232	5.21*	5.51	800	4.41*	5.34
		12	306	4.94*	5.56	967	4.14*	4.49
	Girls	6	113	4.55	3.72	326	5.08	4.78
		7	136	4.19	4.12	3658	4.88	4.67
		8	179	5.26	4.80	364	4.62	4.29
		9	201	4.96	4.46	1354	5.20	5.27
		10	238	5.30	5.54	1545	5.07	5.24
		11	273	5.00	4.81	868	4.56	5.06
		12	296	5.40*	5.34	1007	4.58*	5.37

Note: *N* = number of observations; Std = standard deviation; * = significant twin-singleton mean difference ($p < .05$).
Due to the longitudinal design, the *N*s do not add up to the total number of children.

Table 2

LGM Model Fit Statistics, Model Results with Standard Errors, and the Influence of Sex and SES on the Growth Factors

	Externalizing		Internalizing	
Chi-square value	133.777		171.817	
Degrees of freedom	63		57	
RMSEA	.014		.019	
CFI	.99		.98	
	Singletons	Twins	Singletons	Twins
Intercept mean	8.49 (.25)	8.00 (.08)	4.35 (.20)	4.35 (.08)
Intercept variance	39.75 (3.29)	35.91 (1.41)	12.55 (1.58)*	17.76 (.89)*
Linear slope mean	–.36 (.05)	–.33 (.02)	.41 (.12)	.42 (.06)
Quadr. slope mean	n.s.	n.s.	–.04 (.02)*	–.07 (.01)*
Slope variance	.69 (.12)	.42 (.08)	.32 (.08)*	.51 (.05)*
	Singletons	Twins	Singletons	Twins
Intercept on sex	–3.17 (.51)	–2.29 (.16)	n.s.	.62 (.15)
Slope on sex	n.s.	n.s.	n.s.	n.s.
Quad. slope on sex			n.s.	n.s.
Intercept on SES	–1.52 (.31)	–1.02 (.11)	n.s.	n.s.
Slope on SES	n.s.	n.s.	n.s.	n.s.
Quad. Slope on SES			n.s.	n.s.

Note: CFI = comparative fit index; RMSEA = root mean square error of approximation; n.s. = not significant;
* = significant twin-singleton difference (p <.05).

and singletons. Figure 1 shows the latent growth curves of externalizing problems for twins and singletons, corrected for sex and SES. It shows a linear decrease in externalizing problems over time. The quadratic slope was not significant for both twins and singletons, and was therefore excluded from the model. The differences between the intercept means and variances of twins and singletons were not significant ($\Delta\chi^2(1) = 3.345$, $p = .07$; $\Delta\chi^2(1) = 1.194$, $p = .27$). The linear slope means and variances were also not significantly different between twins and singletons ($\Delta\chi^2(1) = .388$, $p = .53$; $\Delta\chi^2(1) = 3.831$, $p = .05$).

The model fit statistics and model results for internalizing problems are reported in Table 2. The model fit was very good, with CFI = .98 and RMSEA = .019. With regard to the singletons, sex and SES had no significant influence on either the intercept or the slope estimates. For the twins, there was only a significant influence of sex on the intercept ($p < .001$), which indicated that girls had more internalizing problems than boys. The effect sizes for sex differences in the intercepts of internalizing problems were small in both samples (Cohen's d singletons/twins = .16/.15). Figure 2 shows the latent growth curves of internalizing problems of twins and singletons, corrected for SES and sex. It shows a curvilinear increase of internalizing problems in singletons, with levels of problems stabilizing by the age of 12, whereas twins' levels of internalizing problems start to decrease from the age of 9 onwards, resulting in significantly less internalizing problems than singletons by the age of 12. The intercept means of twins and singletons were identical. The development with age, as represented by the linear and quadratic slope means, was different between twins and singletons ($\Delta\chi^2(2) = 20.97$, $p < .001$). Also, twins had significantly larger intercept and slope variances than singletons ($\Delta\chi^2(1) = 7.586$, $p = .01$; $\Delta\chi^2(1) = 4.100$, $p = .04$).

Discussion

This is the first study that longitudinally examined twin-singleton differences in the development of externalizing and internalizing problems in children aged 6 to 12 years. Using latent growth modeling, we estimated and compared growth curves of externalizing and internalizing problems for twins and singletons. For externalizing problems, the growth curves of twins and singletons were similar, showing a linear decrease over time. Decreasing levels of externalizing problems have consistently been found in earlier studies (e.g. Bongers et al., 2003; Leve et al., 2005; Miner & Clarke-Stewart, 2008; Stanger et al., 1997). Although twins showed a tendency to have less externalizing problems than singletons overall, this difference was not significant. We also observed a trend toward a significant variance difference in the slopes of twins and singletons, with singletons having a slightly larger slope variance than twins. Because of the large sample size and the fact that these effects did not reach significance, we conclude that twins are broadly comparable to singletons with regard to the development of externalizing problems. This finding supports the generalizability of twin studies with regard to externalizing problems, and is in line with some previous cross-sectional studies that reported similar levels of externalizing problems for

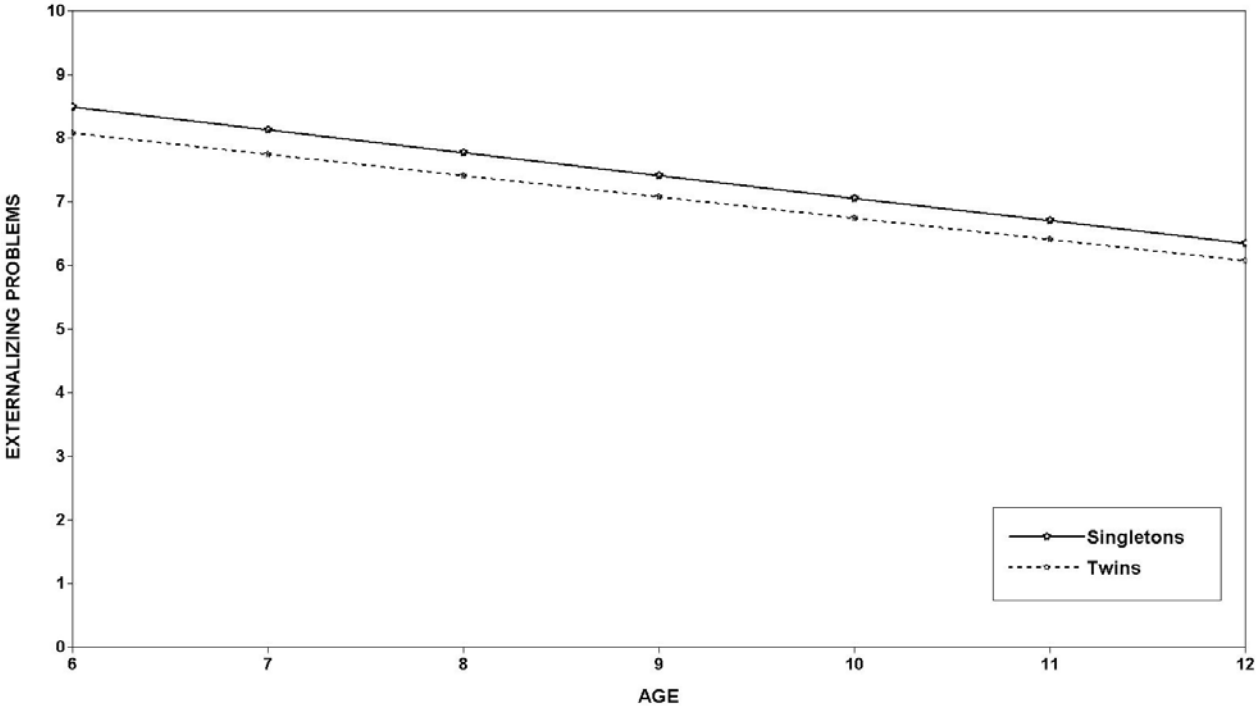

Figure 1

Developmental trajectories of externalizing problems for singletons and twins, adjusted for sex and SES.

twins and singletons (Gjone & Novik, 1995; Pulkkinen et al., 2003).

Internalizing problems seem to develop similarly in twins and singletons up to the age of 9, after which twins' levels of internalizing problems start to decrease in comparison to those of singletons. The lower levels of internalizing problems in twins are in accordance with the results of a previous twin-singleton comparison (Gjone & Novik, 1995). A similar trend of decreasing levels of internalizing problems after the age of 9 has been found earlier (Leve et al., 2005). The interpretation of the larger intercept and slope variances for twins requires some caution. A general problem that may occur when mothers of twins are asked to rate their children, is that they may compare the twins' behavior. The behavior of one twin could then become the standard against which the behavior of the co-twin is rated. This is called a rater contrast effect (Eaves, 1976). Contrast effects may also result from sibling interaction, including cooperation or competition effects. These effects can result in variance differences between twins and singletons, and between monozygotic and dizygotic twins (Bartels et al., 2007). It seems however unlikely that contrast effects have influenced the results, because both externalizing and internalizing problems are measured with the same questionnaire, and we did not find significant variance differences for externalizing problems. Moreover, variances for monozygotic and dizygotic twins did not give any indication to the presence of contrast effects. It also seems unlikely that the difference in sample size between twins and singletons has influenced our findings, because we did not find significant variance differences for externalizing problems, and the variances remained similarly large in smaller random twin samples.

Our results did not support the hypothesis that twins' increased pre- and perinatal risks lead to higher levels of problem behaviors in twins than in singletons. This is probably because the psychological risks associated with these factors are small, and most twins are physically healthy individuals who grow up under normal circumstances (Rutter & Redshaw, 1991). The tendency of twins to have fewer internalizing problems than singletons, might be explained by the fact that twins always have someone close for support. Siblings have been found to be a source of support to each other (Furman & Buhrmester, 1985) and higher levels of sibling support have been associated with lower levels of internalizing problems (Branje et al., 2004). Although many singletons are likely to have a sibling as well, we hypothesize that the presumed effect of sibling support is stronger for twins than for regular siblings, since twins may have a more intimate relationship with each other than regular siblings (Segal et al., 2008). Consequently, twinship may be a protective factor to the development of internalizing problems in early adolescence. Environmental influences that are shared by twin pairs and not by other members of the family may have an effect. Twin researchers should therefore use extended twin designs that include non-

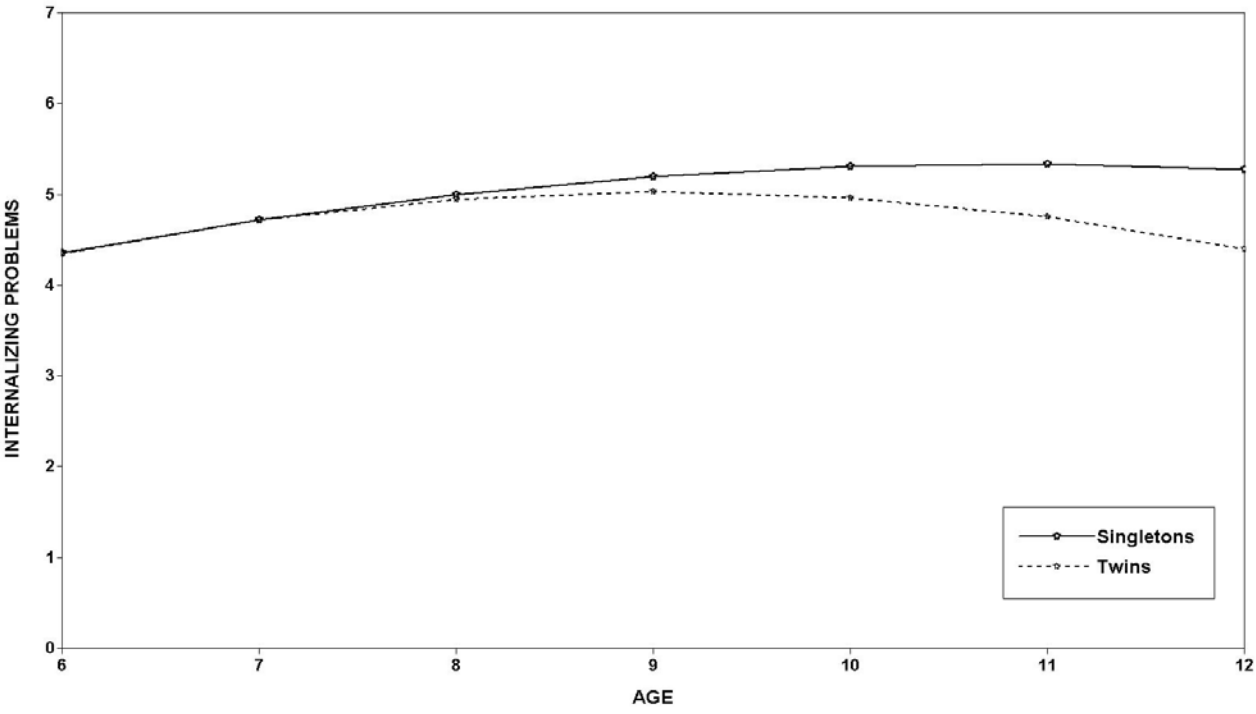

Figure 2

Developmental trajectories of internalizing problems for singletons and twins, adjusted for sex and SES.

twin siblings, and perform explicit tests of a special twin environment for internalizing problems.

Differences between our findings and those from earlier studies could be due to the different designs (cross-sectional versus longitudinal), different age groups, and the use of different measures. All previously conducted studies were cross-sectional. This study provides a clear picture of twin-singleton differences in the development of externalizing and internalizing problems in childhood, in addition to cross-sectional differences. Finally, we acknowledge that examining growth curves in addition to cross-sectional differences is not sufficient to fully evaluate the generalizability of twin studies with regard to these types of behaviors, but it has brought us a step further in this rather unexplored field of research.

Several limitations of the present study need to be considered. First, reorganization of the data matrix as a function of chronological age created a missing data problem. Although we chose a statistical approach that can properly handle missing data, it would have been better if we had data of each respondent at each age. Second, we only used maternal ratings. Unfortunately, there were not enough father ratings available in the singleton sample for a reliable twin-singleton comparison. Future research should also include father ratings or teacher ratings to obtain a more complete picture of the behavior of the child, accounting for example for situational specific behavior. Third, the two samples are not from the same time periods and were recruited from different regions. An earlier study that used the same singleton data did not find evidence for clear secular differences in psychopathology over a 10-year period (1983–1993; Verhulst et al., 1997). Tick et al. (2007) found increases in Dutch children's problems over a 20-year period (1983–2003), but these increases were not consistent across informants and across age. Only parent reports for children aged 6 to 16 years showed a very small increase in internalizing problems, but this increase was not reported for self-reports and teacher reports for children of the same age and not for parent reported problems in preschool children. With regard to our twin sample, we did not find any significant differences in the mean externalizing and internalizing problem scores at ages 7, 10, and 12 between twins from different cohorts. Considering these findings, it seems unlikely that cohort effects can explain twin-singleton differences. The different regions the samples have been recruited from does not limit the comparability of the samples, since we found that twins from the province of Zuid-Holland have the same externalizing and internalizing trajectories as twins from the rest of the Netherlands. Also, Tick et al. (2007) showed that there were no significant differences in mean scale scores on the CBCL between children living in Zuid-Holland and children living elsewhere in the Netherlands.

In conclusion, our findings confirm the generalizability of twin studies with regard to the development of externalizing problems in middle and late childhood, since developmental trajectories of externalizing problems were similar for twins versus singletons.

However, the generalizability of studies on internalizing problems in early adolescence in twin samples should be addressed with care. Twinship may be a protective factor for the development of internalizing problems during early adolescence. Our findings regarding internalizing problems are indicative of twin-singleton differences in adolescence. Future research should extend our findings by describing developmental trajectories of externalizing and internalizing problems from childhood to adulthood.

Acknowledgments

Data collection was funded by the following grants: 'Genetics of externalizing disorders in children' (NWO 904-57-94), 'Spinozapremie' (NWO/SPI 56-464-14192), 'Twin-family database for behavior genetics and genomics studies' (NWO 480-04-004), 'Twin research focusing on behavior and depression' (NWO 400-05-717), 'Developmental Study of Attention Problems in Young Twins' (NIMH, RO1 MH58799-03), and 'Bridge Award' (NIMH R56). Sylvana Robbers was supported by the Sophia Children's Hospital Foundation, project number 483.

Abbreviations

CBCL: Child Behavior Checklist; LGM: latent growth modeling; CFI: comparative fit index; RMSEA: root mean square error of approximation.

References

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

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

Bartels, M., Boomsma, D. I., Van Beijsterveldt, C. E. M., Hudziak, J. J., & Van den Oord, E. J. (2007). Twins and the study of rater (dis)agreement. *Psychological Methods, 12*, 451–466.

Bartels, M., Van Beijsterveldt, C. E. M., Derks, E. M., Stroet, T. M., Polderman, T. J. C., Hudziak, J. J., & Boomsma, D. I. (2007). Young Netherlands Twin Register (Y-NTR): A longitudinal multiple informant study of problem behavior. *Twin Research and Human Genetics, 10*, 3–11.

Bleker, O. P, Oosting, J., & Hemrika, D. J. (1988). On the cause of the retardation of fetal growth in multiple gestations. *Acta Geneticae Medicae et Gemellologiae (Roma), 37*, 41–46.

Bongers, I. L., Koot, H. M., Van der Ende, J., & Verhulst, F. C. (2003). The normative development of child and adolescent problem behavior. *Journal of Abnormal Psychology, 112*, 179–192.

Boomsma, D. I., De Geus, E. J., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J., Posthuma, D., van Beijsterveldt, T. C., Hudziak, J. J., Bartels, M., & Willemsen, G. (2006). Netherlands Twin Register: From twins to twin families. *Twin Research and Human Genetics, 9*, 849–857.

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.

Boomsma, D. I., Vink, J. M., Van Beijsterveldt, T. C. E. M., De Geus, E. J., Beem, L., Mulder, E. J. C. M., et al. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research, 5*, 401–406.

Branje, S. J. T., Van Lieshout, C. F. M., Van Aken, M. A. G., & Haselager, G. J. T. (2004). Perceived support in sibling relationships and adolescent adjustment. *Journal of Child Psychology and Psychiatry, 45*, 1385–1396.

CBS. (2001). *Standard Classification of Occupations*. Heerlen/Voorburg: Central Bureau of Statistics.

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

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology, 30*, 415–441.

Estourgie-van Burk, G. F., Bartels, M., van Beijsterveldt, T. C., Delemarre-van de Waal, H. A., & Boomsma, D. I. (2006). Body size in five-year-old twins: Heritability and comparison to singleton standards. *Twin Research and Human Genetics, 9*, 646–655.

Furman, W., & Buhrmester, D. (1985). Children's perceptions of the personal relationships in their social networks. *Developmental Psychology, 21*, 1016–1024.

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.

Gjone, H., & Novik, S. (1995). Parental ratings of behaviour problems: A twin and general population comparison. *Journal of Child Psychology and Psychiatry, 36*, 1213–1224.

Gruenwald, P. (1970). Environmental influences on twins apparent at birth. A preliminary study. *Biology of the Neonate, 15*, 79–93.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: Externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology, 28*, 161–179.

Lahti, J., Räikkönen, K., Kajantie, E., Heinonen, K., Pesonen, A. K., Järvenpää, A. L., et al. (2006). Small body size at birth and behavioral symptoms of ADHD in children aged five to six years. *Journal of Child Psychology and Psychiatry, 47*, 1167–1174.

Leve, L. D., Kim, H. K., & Pears, K. C. (2005). Childhood temperament and family environment as predictors of internalizing and externalizing trajectories from ages 5 to 17. *Journal of Abnormal Child Psychology, 33*, 505–520.

Levy, F., Hay, D., McLaughlin, M., Wood, C., & Waldman, I. (1996). Twin-sibling differences in parental reports of ADHD, speech, reading and behaviour problems. *Journal of Child Psychology and Psychiatry, 37*, 569–578.

McArdle, J. J., & Erpstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110–133.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*, 185–204.

Miner, J. L., & Clarke-Stewart, K. A. (2008). Trajectories of externalizing behavior from age 2 to age 9: relations with gender, temperament, ethnicity, parenting, and rater. *Developmental Psychology, 44*, 771–786.

Moilanen, I., Linna, S. L., Ebeling, H., Kumpulainen, K., Tamminen, T., Piha, J., et al. (1999). Are twins' behavioural/emotional problems different from singletons'? *European Child & Adolescent Psychiatry, 8*, 62–67.

Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus User's Guide. Fifth Edition.* Los Angeles, CA: Muthén & Muthén.

Phillips, D. I., Davies, M. J., & Robinson, J. S. (2001). Fetal growth and the fetal origins hypothesis in twins – Problems and perspectives. *Twin Research, 4*, 327–331.

Pulkkinen, L., Vaalamo, I., Hietala, R., Kaprio, J., & Rose, R. J. (2003). Peer reports of adaptive behavior in twins and singletons: Is twinship a risk or an advantage? *Twin Research, 6*, 106–118.

Raikkonen, K., Pesonen, A. K., Heinonen, K., Kajantie, E., Hovi, P., Jarvenpaa, A. L., et al. (2008). Depression in young adults with very low birthweight: The Helsinki study of very low-birthweight adults. *Archives of General Psychiatry, 65*, 290–306.

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.

Segal, N. L., Chavarria, K. A., & Hoven Stohs, J. (2008). De leefwereld van tweelingen. In D. I. Boomsma (Ed.), *Tweelingonderzoek. Wat meerlingen vertellen over de mens*. Amsterdam: VU Uitgeverij.

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

Stanger, C., Achenbach, T. M., & Verhulst, F. C. (1997). Accelerated longitudinal comparisons of aggressive versus delinquent syndromes. *Development and Psychopathology, 9*, 43–58.

Tick, N. T., Van der Ende, J., & Verhulst, F. C. (2007). Twenty-year trends in emotional and behavioral problems in Dutch children in a changing society. *Acta Psychiatrica Scandinavica, 116*, 473–482.

Van den Oord, E. J., 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 Westerlaak, J. H., Kropman, J. A., & Collaris, J. W. M. (1975). *Beroepenklapper (Manual for occupational level)*. Nijmegen: Instituut voor Sociologie.

Verhulst, F. C., Akkerhuis, G. W., & Althaus, M. (1985). Mental health in Dutch children: A cross-cultural comparison. *Acta Psychiatrica Scandinavica, 323*, 1–108.

Verhulst, F. C., Van der Ende, J., & Koot, H. M. (1996). *Handleiding voor de CBCL/4–18 [Dutch manual for the CBCL/4-18]*. Rotterdam, the Netherlands: Academic Medical Center Rotterdam / Erasmus University, Sophia Children's Hospital, Department of Child and Adolescent Psychiatry.

Verhulst, F. C., Van der Ende, J., & Rietbergen, A. (1997). Ten-year time trends of psychopathology in Dutch children and adolescents: No evidence for strong trends. *Acta Psychiatrica Scandinavica, 96*, 7–13.