A Comparison of Educational Achievement in a National Sample of Dutch Female Twins and Their Matched Singleton Controls

Celina C.C. Cohen[1], Stephanie H.M. van Goozen[2], Jacob F. Orlebeke[3], Jan K. Buitelaar[1], and Peggy T. Cohen-Kettenis[1]

[1]Department of Child and Adolescent Psychiatry, University Medical Centre Utrecht, and the Rudolf Magnus Institute for Neurosciences, The Netherlands

[2] Department of Psychiatry, University of Cambridge, Cambridge, UK

[3] Department of Biological Psychology, Vrije Universiteit Amsterdam, The Netherlands

The aim of this study was to compare the educational achievement of female twins (*n* = 577) and their matched singleton controls (*n* = 447), who were selected from participants of a national test of educational achievement in the years 1993 to 1998. To assure the representativeness of the selected groups we also compared the achievement scores of the twins and the controls to those of the total Dutch female population tested in the same period. We analyzed the results of the following educational achievement scales: Language, Mathematics and Information Processing. The results indicated that the singleton classmates performed significantly better than the twins on all three scales. However, the twins performed equally well as compared to the Dutch female population. We believe that our singleton control group was not as properly selected as we intended, a selection bias operative at the level of the schools may have confounded the comparison. We therefore conclude on the basis of a comparison with the performance of the total Dutch female population that there are no differences in educational achievement between female twins and singletons.

In research on intellectual functioning of twins the selection of a proper control is a complicated issue. In twin research control groups often consist of singletons, but it is not clear if singletons and twins can be compared directly for several reasons.

Since twins weigh on average about 1000 grams less than singletons, and low birth weight is statistically associated with lower intelligence and other less favorable developmental and health outcomes during childhood and later life (Barker et al., 1992; McCormick et al., 1990; Seidman et al., 1992; Veen et al., 1991), it is to be expected that intelligence and associated school performances are lower in twins than in singletons. In addition to birth weight, other aspects of the twin condition such as intra-uterine growth retardation and shorter gestational length (Powers & Kiely, 1994) may be responsible for a twin–singleton difference in intellectual performance (Posthuma et al., 2000). Several studies have shown that mean intelligence scores are lower for twins than for singletons (Record, McKeown, & Edwards, 1970). However, in a recent study Posthuma et al. (2000) showed that the selection criteria for the composition of a singleton control group may be essential for the outcome of the study. They selected what has been called the perfect control group for a group of twins namely, their non-twin siblings. They showed that by using these subjects differences in intelligence between twins and their non-twin siblings disappeared. Indeed, in order to control for various confounding socioeconomic and genetic factors the use of family members seems ideal. However, this option is not always possible, for example when studies have to control for variables such as age and sex. In these circumstances the next best option is the selection of singleton controls from different but similar rearing environments that match as closely as possible to the twin. Important matching criteria are sex, age, social economical status and rural area. For example, Nathan and Guttman (1984) compared twins and singletons with respect to cognitive functioning whom were reared in the same kibbutz society in Israel. Although they found differences in cognitive performance between twins and singletons they did not ascribe these to the fact of being a twin because it turned out that the mothers of the twins had relatively fewer years of education. While the kibbutz community offers the unique opportunity to control for some relevant confounding factors, for obvious reasons not many of these studies have been conducted. In our western societies the selection of a classmate could be considered to be a relatively good control case. Since it is known that schools have an important influence on the development of intellectual abilities and pupils attending the same school usually share a similar social economical background, and can easily be matched on age and sex, we considered singletons coming from the same classroom as the twins to be proper controls.

We conducted a nation-wide study on educational achievement of Dutch female twins and their female classroom singleton controls. Moreover, to take the effect of

Address of correspondences: Celina Cohen M.Sc, Department of Child and Adolescent Psychiatry, A01.468, University Medical Centre Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands. Email: c.cohen@psych.azu.nl

Celina C.C. Cohen, Stephanie H.M. van Goozen, Jacob F. Orlebeke, Jan K. Buitelaar, and Peggy T. Cohen-Kettenis

the domestic environment into account, the matched control girl had to have an older sister in case the twin couple consisted of two girls or she had to have an older brother in case the twin couple consisted of opposite sex children. The focus of the study was on girls only because the present study is a part of a larger prospective follow-up project on female twins.

The aim of the study was to examine whether the educational achievement of female twins and their matched singleton controls, who were selected from participants to a national test of educational achievement in the years 1993 to 1998, would be similar. To assure the representativeness of the selected groups we also compared the achievement scores of the twins and the controls to those of the total Dutch female population tested in the same period.

Materials and Methods

Procedure

The study population consisted of a nation-wide sample of 6th grade participants that participated in the "CITO" elementary test (a national test of educational achievement; Citogroep, 1997) between the years 1993–1998. The subjects were girls with a male or female co-twin and singleton girls with an older brother or older sister. In order to protect the anonymity of the participants the CITO organization does not register the date of birth of the participants. To find out whether schools, who had taken part in the CITO test, had participating twins it was decided to search the database for double surnames in the same school and year. Schools with at least six double surnames were contacted by letter. These turned out to be 742 schools in the Netherlands and all schools were approached by mail.

In a letter the school was asked if the names on the list belonged to a twin pair and that if this was indeed the case, whether they would be prepared to give the names of the twin pair and the name of a singleton female classmate — with an older brother or sister.

Sample

Using this procedure 222 girls with a boy co-twin and 359 girls with a girl co-twin were selected as participants to the study. We could not determine zygosity of the twins because we did not have access to blood-samples of the twins since they were anonymous subjects to us. The data collection was conducted through the CITO organization and schools.

For comparison, 212 singleton girls with an older brother and 238 singleton girls with an older sister were selected. All participants were between 11 and 12 years old.

Measures

Educational achievement was assessed by the Dutch CITO-elementary test. The CITO consists of 240 multiple-choice items assessing four different intellectual skills: Language, Mathematics, Information Processing, and World Orientation. Each performance scale contains 5 or 6 sub-scales, with a total of 60 multiple-choice questions. Together these performance scales result in a standardized score between 501 and 550. The test is administered on three consecutive days.

Because we used the data of the years 1993 to 1998 and the World Orientation scale was only added to the CITO-test after 1994, we analyzed the results of the performance scales Language, Mathematics and Information Processing only.

Statistics

The Statistical Package for Social Science (SPSS) was used to analyze the data.

Independent samples t tests were used to compare the mean scores of the twin and singleton groups.

In order to examine whether the results of our twin and singleton groups were representative to those of the total Dutch female population we also tested whether there were differences in educational achievement between these three groups. One sample t tests were used to analyze these differences.

The magnitude of the effect size was calculated according to Cohen (1988). Effect size is the difference between two means, divided by the pooled standard deviation of the two groups (twins and singletons). In general, an effect size of approximately 0.80 is considered large, 0.50 moderate, and 0.20 small (Cohen, 1988).

The significance level for all analyses was set at $p < 0.05$.

Results

Twins and Singletons

Within the twin group no significant differences were found between same-sex and opposite-sex twin girls. Nor were significant differences found within the singleton group between girls with an older sister or brother. We therefore combined these subgroups into a twin and singleton group.

When we compared the total mean CITO-test scores over the years 1993–1998 of singletons and twins a significant group difference was found on the mean CITO total score with the singleton group outperforming the twin group: Mean Singletons = 537.47 (SD = 8.93), Mean Twins = 535.45 (SD = 9.61), t = 3.43, $p < .001$.

We also analyzed whether a group difference was found for the three separate performance scales. Table 1 shows the mean performance of the twins and singletons on the three main scales of the CITO-test. The results clearly indicate that the singletons performed significantly better than the twins on all three performance scales (see Table 1).

Finally, in order to find out whether the scores of our twin and singleton groups were representative to those of the total Dutch female population we compared the mean scores of these three groups for each performance scale. In Table 1 the mean scores for each group are presented.

When we compared the mean scores of the singleton and twin groups to those of the Dutch female population we found that the data indicated that the singleton group performed significantly better than the Dutch female population on all three achievement scales. Whereas there were no differences in performance between the twin group and the Dutch female population.

The effect sizes found between the twins and singletons for the three educational achievement scales were respectively for, Language 0.21, Mathematics 0.20, and for Information Processing 0.18. Very small effect sizes were found between

Table 1

Mean CITO Scale Scores (1993–1998) of the Total Dutch Female Population, the Female Singleton Controls and the Female Twins are Presented Together with the Standard Deviation (The Significance of the Differences is Indicated)

	Population (P) $n = 340877$	Singletons (S) $n = 447$	Twins (T) $N = 577$	t(SXT)	t(PXT)	t(SXP)
Language	43.22(9.22)	45.29 (8.57)	43.44 (8.99)	3.32**	0.59	5.10**
Mathmatics	40.57(11.46)	43.31 (10.82)	41.06 (11.35)	3.21**	1.03	5.35**
Information Processing	42.93(9.59)	44.81 (8.64)	43.19 (9.47)	2.87*	0.65	4.61**
Total		537.47 (8.93)	535.45 (9.61)	3.43**		

Note: $*$ = $p < 0.01$, $**$ = $p < 0.001$

the twins and the total Dutch female population respectively, 0.02, 0.04, and 0.03.

Discussion

It is still unclear whether twins function intellectually more poorly than singletons. For example, Record et al. (1970) found lower scores for twins compared to singletons that could not be explained by factors like birth weight and gestational length. On the other hand, Posthuma et al. (2000) showed a similar intellectual functioning of twins compared to their non-twin siblings. An important explanation for these mixed findings could be related to the composition of the control groups for comparison.

The aim of the present study was to investigate whether female twins and their best possible selected singleton controls, matched for sex, age, school, rural area and domestic environment, would be similar in educational achievement. To this end, the results on 6 consecutive birth cohort years of the CITO-test, a nation-wide test on educational achievement, were used.

The results indicate that the female twin group (n = 577) performed significantly worse on all three scales of the CITO-test compared to their matched singleton female controls (n = 447). On the basis of these results one could conclude that the educational achievement of twins is indeed worse than that of their matched singleton controls. This conclusion is in line with other studies and would imply that twins do have a disadvantage in this respect. Studies have shown that, for example, low birth weight has a negative impact on cognitive development (Seidman et al., 1992) and it is known that twins are generally born with a lower mean birth weight than singletons.

However, despite the fact that we tried to compose the best possible singleton control group the selection of the control cases by the participating schools may have had an unforeseen consequence. It is possible that our singleton control group was not as properly selected as we intended. The statistical analyses comparing the female twin data to those of female population showed that the twin means were more similar to the population means than to the female singleton means: the scores of our singleton control group were clearly higher than those of the populations means. It is possible that this is due to a selection bias in the choice of the singleton controls by the schools participating in the study. The fact that the study involved a national comparison of female CITO test scores, and that the CITO outcomes are used to compare the performances of schools,

could have influenced the selection procedure of the teachers in them choosing intellectually well-performing female controls as comparison subjects. If we examine the results of the comparisons of scores of the twins and the female national population it becomes clear that these are similar and we therefore could also conclude that there are no differences in educational achievement between female twins and singletons. The conclusion is supported by the negligible effect sizes we found.

Although we intended in this study to select the best possible control group of singletons by selecting classmates of the twins, a selection bias operative at the level of the schools may have confounded the comparison. In the future this may be overcome by an at random selection of control cases from all possible candidates that match the selection criteria.

Acknowledgments

References

Barker, D. J. P., Osmond, C., Golding, J., Kuh, D., & Wadsworth, M. E. J. (1992). Growth utero, bloodpressure in childhood and adult life, and mortality from cardiovascular disease. In D. J. P. Barker (Ed.), *Fetal and Infant Origins of Adult Disease* (pp. 165–174). London: British Medical Journal.

Citogroep. (1997). *Eindtoets Basisonderwijs 1997*. Arnhem: Citogroep.

Cohen, J. (1988). *Statistical power analyses for behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

McCormick, M. C., Gortmaker, S. L., & Sobol, A. M. (1990). Very low birth weight children: Behavior problems and school difficulty in a national sample. *Journal of Pediatrics, 117,* 687–693.

Nathan, M., & Guttman, R. (1984). Similarities in test scores and profiles of kibbutz twins and singletons. *Acta Geneticae Medicae et Gemellologiae Roma, 33,* 213–218.

Posthuma, D., De Geus, E. J. C., Bleichrodt, N., & Boomsma, D. I. (2000). Twin-singleton differences in intelligence? *Twin Research, 3,* 83–87.

Powers, W. F., & Kiely, J. L. (1994). The risks confronting twins: A national perspective. *American Journal of Obstetrics and Gynaecology, 170,* 456–461.

Record, R. G., McKeown, T., & Edwards, J. H. (1970). An investigation of the difference in measured intelligence between twins and single births. *Annals of Human Genetics, 34,* 11–20.

Seidman, D. S., Laor, A., Gale, R., Stevenson, D. K., Mashiac, S., & Danon, Y. L. (1992). Birth weight and intellectual performance in late adolescence. *Obstetrics and Gynecology*, *79,* 543–546.

Veen, S., Ens-Dokkum, M. H., Schreuder, A. M., Verloove-Vanhorick, S. P., Brand, R., & Ruys, J. H. (1991). Impairments, disabilities, and handicaps of very preterm and very-low-birthweight infants at five years of age. *Lancet, 338,* 33–36.