Twin–singleton differences in intelligence?

Daniëlle Posthuma1, Eco JC De Geus1, Nico Bleichrodt2 and Dorret I Boomsma1

1Department of Biological Psychology and 2Department of Work and Organizational Psychology, Vrije Universiteit Amsterdam, The Netherlands

The twin method has been criticised for its alleged non-generalisability. When population parameters of intellectual abilities are estimated from a twin sample, critics point to the twin–singleton differences in intrauterine and family environments. These differences are suggested to lead to suboptimal cognitive development in twins. Although previous studies have reported twin–singleton differences in intelligence, these studies had two major drawbacks: they tested young twins, and twins were compared with (genetically) unrelated singletons. To test accurately whether twin–singleton differences in intelligence exist, a group of adult twins and their non-twin siblings were administered the Dutch WAIS-III. The group was large enough to detect twin–singleton differences of magnitudes reported in earlier investigations. The data were analysed using maximum likelihood model fitting. No evidence of differences between adult twins and their non-twin siblings on cognitive performance was found. It is concluded that twin studies provide reliable estimates of heritabilities of intellectual abilities which can be generalised to the singleton population. Twin Research (2000) 3, 83–87.

Keywords: twin study, intelligence, twins, singletons

Introduction

Classic behavioural genetic studies provide statistical estimates of heritabilities that form the first step in the search for genes for complex behaviour.1,2 A large part of these behavioural genetic studies are based on twin samples. These samples have sometimes been criticised for their alleged non-generalisability; since twins are ‘special’ they may not be representative of singletons. Especially in the field of cognitive abilities twins are generally considered to be at a disadvantage compared with singletons.3–6

Twins share the womb at the same time and consequently share prenatal nutrition provided by the mother’s dietary intake. When preparing for labour, twins compete for the best position. This suboptimal intrauterine environment may lead to prematurity, low birth weight and lower weight-for-gestational age,7 which in turn in several cases have been associated with low childhood IQ.8–12 Apart from a general suboptimal intrauterine environment for both twins, it is known that one of the two foetuses will suffer more from this suboptimal environment than the other.13 It is usually the second-born twin that experiences the greatest adverse effects of sharing the womb.14

Besides these adverse effects of sharing the womb twins may suffer from twin-related stresses in the family environment in which they are reared. A multiple birth puts stress on a family which may have a negative effect on the (cognitive) development of a twin pair. In some studies it is argued that especially for monozygotic (MZ) twins, who are very much alike, limitation of resources and competition may lead to negative influences for at least one twin member.3

A relatively small number of studies has been devoted to detecting twin–singleton differences in cognition.4,6,15 The one study that stands out was conducted by Record, McKeown and Edwards6 who compared an impressive number of singletons, twins and even a few triplets. Verbal reasoning scores from the British eleven-plus examination were gathered from 48 913 singletons, 1082 twin pairs and eleven triplets. Standard verbal reasoning scores were significantly lower for twins (standard verbal IQ 95.7) than for singletons (100.1). Triplets performed even worse (91.6). The authors investigated whether this 4.4 standard points difference between twins and singletons could be attributed to effects of maternal age, birth weight, gestational age, zygosity and whether a twin was born first or second. None of these factors could explain the difference.

Record et al15 also investigated whether twins of whom one had died shortly after birth differed from singletons; although for these ‘twins’ a slightly lower score than normal singletons (1.9 points) was found, this difference was much smaller than the 4.4 points difference between singletons and twins of which both members were still alive. Based on this observation the authors concluded that the difference of

Editor: J. Heath

4.4 points between singletons and twins cannot be attributed to negative effects of sharing the womb, but instead must be sought in the environment in which twins are reared. However, since Record et al. did not control for any difference in twin families and singleton families, they could not rule out selection biases in the sampling of twin and non-twin families. Such biases may exist because twins as a group may have a slightly different genetic or social background than singletons.

Nathan and Gutman tried to overcome selection bias in twin and singleton families by comparing twins and singletons (aged 8–13 years) who were reared in the same kibbutz. A kibbutz is an Israeli community in which children are collectively reared. So although the twins and singletons in this study did not have the same genetic background, they were accurately matched for family environment and childrearing practices. In this study dizygotic (DZ) twins performed worse than MZ twins and singletons. According to the authors, however, this difference could be totally ascribed to the relatively few years of schooling of the group of DZ mothers. Thus, in spite of the attempt to match twins and singletons this study is also an example of biased family sampling.

In addition to comparing twins with familially unrelated singletons, most previous studies have been conducted using young twins. Because these studies show that twins recover any deficits in intellectual performance by 6–8 years of age, the comparison of twins and singletons at ages below 8 years does not provide a good indication of adult twin–singleton differences. To the best of our knowledge studies comparing the IQ of adult twins and genetically related singletons have not yet been conducted.

In the present study mean scores of adult MZ and DZ twins on intellectual ability are compared with the mean scores of their non-twin siblings. Non-twin siblings make an ideal control group; both genetic background and early familial environments are perfectly matched.

Method

Subjects

The subjects were 358 family members from a total of 152 twin families who participated in a project investigating the genetics of adult brain function. The Dutch version of the Wechsler Adult Intelligence Scale-III (WAIS-III) was administered when the participants visited the laboratory for a combined session of neuropsychological and electroencephalographic measurements. All subjects were recruited from the Netherlands Twin Registry. The twins had previously participated in one of two previously conducted studies in which zygosity was assessed by blood group polymorphisms and DNA typing.

In total, 98 siblings, 101 MZ twins, 153 DZ twins and 9 triplets participated. Since the group of triplets was small, we discarded the data of the last born of the triplets and treated the remaining two members as if they were twins. This left 98 siblings and 260 twins. The study recruited twins pairs at most two of their non-twin siblings. It also included single twins (co-twin refused participation) and siblings only (both twins refused). Thus, families consisted of at least one member and at most four members. Table 1 shows the number of families with a particular constitution, e.g., 27 MZ families consisting of two twin members and no siblings participated; siblings from nine families participated without the twins. Due to administrative errors five individual test scores are missing subtest digit symbol-coding, four individual test scores are missing subtests block design and digit symbol-free recall, and one individual test score is missing subtest digit symbol-pairing and subtest letter-number sequencing. Results are based on the available number of subjects per subtest (see Table 3).

Mean age and sex distribution per group are displayed in Table 2. Of the 98 non-twin siblings, 35 were younger than the twin from the same family, and 63 were older. Distribution of sex did not differ in the DZ twins and the siblings. Slightly fewer female MZ twins than male MZ twins participated.

<table>
<thead>
<tr>
<th>Group</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
<th>Mean age in years (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ twins</td>
<td>58</td>
<td>43</td>
<td>101</td>
<td>39.7 (12.63)</td>
</tr>
<tr>
<td>DZ twins</td>
<td>70</td>
<td>89</td>
<td>159</td>
<td>37.3 (11.87)</td>
</tr>
<tr>
<td>Sibs</td>
<td>46</td>
<td>52</td>
<td>98</td>
<td>37.1 (12.02)</td>
</tr>
</tbody>
</table>

sd = standard deviation

Table 1 Sample configuration

<table>
<thead>
<tr>
<th>number of non-twin siblings</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ twins</td>
<td>2 twins</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td>DZ twins</td>
<td>2 twins</td>
<td>32</td>
<td>27</td>
</tr>
<tr>
<td>No twin</td>
<td>1 twin</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Total non-twin siblings</td>
<td>-</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

Total: 98
Procedure

Eleven subtests of the Dutch WAIS-III were administered in a fixed order. Subtests included block design, letter-number sequencing, information, matrix reasoning, similarities, picture completion, arithmetic, vocabulary, digit symbol coding, digit symbol pairing and digit symbol free recall. Age and sex normalised scores for the Dutch WAIS-III are not yet available; raw scores were used in the analyses throughout. All subjects were paid Dfl. 50. - for participation.

Statistical analyses

As can be seen from Table 1 the data were characterised by the varying number of participating family members; families consisted of one to four members which could be any combination of one or two twins and/or non-twin siblings. This variability in number of observations per family causes serious computational problems. In Mx24 the handling of such 'incomplete' data is implemented by calculating twice the negative log-likelihood (–2LL) of the raw data for each family, with the following formula:

\[ –2LL = -k \log (2\pi) + \log |\Sigma| + (x_i - \mu_i)'\Sigma^{-1}(x_i - \mu_i), \]

where \( k \) (\( k = 1, 2, 3 \) or \( 4 \)) denotes the number of observed variables within a family, \( \Sigma \) \((4 \times 4)\) is the covariance matrix of family members, \( x_i \) (for \( i = 1, 2, 3, 4 \)) is the vector of observed scores, \( \mu_i \) is the column vector of the estimated means of the variables, and |\( \Sigma| \) and \( \Sigma^{-1} \) are the determinant and inverse of matrix \( \Sigma \), respectively.

When two models which provide –2LLs are nested, subtracting the two –2LLs from each other provides a \( \chi^2 \) distribution. A high \( \chi^2 \) against a low gain of degrees of freedom \((\Delta df)\) denotes a worse fit of the second, more restrictive model relative to the first model.

Four univariate nested models were fitted using this procedure. In the first model all means were estimated individually. The second model is the same as the first model with two extra equality constraints; one on the means of both members of the MZ twin pairs and another one on the means of both members of the DZ twin pairs. The third model is the same as the second model but further constrains the means of the MZ twin pairs and the DZ twin pairs to be equal. The fourth is the same as the third model but with an extra equality constraint on the means of all twins (MZ and DZ) and siblings.

Model 2 tests whether the means of first born twins and second born twins within zygosity groups are significantly different. The third model serves as a test of the assumption that the means in MZ twins and DZ twins do not differ. Model 4 tests whether the means of twins and siblings are significantly different.

For all models the variances of all twin members and all siblings were constrained equal, and all covariances of all twin sib pairs, the covariance of two sibs within one family and the covariance of the DZ twins were set equal.

Statistical power

We calculated the necessary sample size for each group (singleton and twins) based on the effect size as found in Record et al's study.6 A measure of effect size that is independent of scaling is Cohen's \( d \), which is calculated as follows:

\[ d = (\mu_1 - \mu_2)/\sigma \]

where \( \mu_1 \) is the mean of the first group (singleton), \( \mu_2 \) is the mean of the second group (twins) and \( \sigma \) is the common standard deviation.25

Record et al6 found a 4.4 standard points difference between the two groups. The standard deviation of an IQ score is by definition 15. The effect size in the Record et al study was thus 0.29, which is considered a small effect. For a one-tailed test with \( \alpha = 0.05, 1 - \beta = 0.80 \), and two related samples, 70 individuals per group (singleton and twins) are needed to detect an effect of such small magnitude.66 We had 260 twins and 98 non-twin siblings giving us the power to detect effect sizes well below 0.29.

Results

The observed means and standard deviations of WAIS-III subtests per group are displayed in Table 3.

<table>
<thead>
<tr>
<th>subtest</th>
<th>mz twin (N = 101)</th>
<th>dz twin (N = 159)</th>
<th>sibs (N = 98)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block design</td>
<td>26.20 (8.96)</td>
<td>25.72 (9.28)</td>
<td>26.25 (8.85)</td>
</tr>
<tr>
<td>LN sequencing</td>
<td>12.21 (3.42)</td>
<td>11.21 (2.61)</td>
<td>11.86 (2.90)</td>
</tr>
<tr>
<td>Information</td>
<td>23.41 (6.32)</td>
<td>23.93 (6.00)</td>
<td>24.11 (6.54)</td>
</tr>
<tr>
<td>Matrix reasoning</td>
<td>19.36 (3.38)</td>
<td>19.16 (3.44)</td>
<td>19.40 (3.28)</td>
</tr>
<tr>
<td>Similarities</td>
<td>26.91 (5.58)</td>
<td>27.17 (5.43)</td>
<td>27.33 (5.58)</td>
</tr>
<tr>
<td>Picture completion</td>
<td>20.86 (2.55)</td>
<td>20.72 (2.60)</td>
<td>20.55 (3.18)</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>13.86 (3.86)</td>
<td>13.75 (3.89)</td>
<td>14.70 (4.12)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>49.07 (11.60)</td>
<td>48.26 (10.55)</td>
<td>47.83 (13.54)</td>
</tr>
<tr>
<td>DS coding</td>
<td>76.09 (15.22)</td>
<td>77.66 (19.52)</td>
<td>78.83 (15.86)</td>
</tr>
<tr>
<td>DS free recall</td>
<td>7.63 (1.20)</td>
<td>7.54 (1.12)</td>
<td>7.54 (1.27)</td>
</tr>
<tr>
<td>DS pairing</td>
<td>13.25 (4.25)</td>
<td>12.67 (4.19)</td>
<td>12.92 (4.02)</td>
</tr>
</tbody>
</table>

Table 3: Observed means and standard deviations of WAIS-III subtests per group

---

\*based on 157 observations
\*based on 96 observations
\*based on 97 observations
\*based on 158 observations
\*based on 94 observations
\*based on 99 observations

LN = Letter-number
DS = Digit symbol
Table 4: Fit indices for nested sequence of models fitted to raw data of WAIS-III subtest scores of MZ twins, DZ twins and siblings

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>–2LL df</td>
<td>–2LL df</td>
<td>–2LL df</td>
<td>–2LL df</td>
<td>χ² (df = 7)</td>
</tr>
<tr>
<td>Block design</td>
<td>2451.48 343</td>
<td>2453.90 345</td>
<td>2454.01 346</td>
<td>2459.56 350</td>
<td>8.08 n.s.</td>
</tr>
<tr>
<td>Letter-number sequencing</td>
<td>1738.22 346</td>
<td>1739.38 348</td>
<td>1744.57 349</td>
<td>1750.37 353</td>
<td>12.15 n.s.</td>
</tr>
<tr>
<td>Information</td>
<td>2194.37 347</td>
<td>2197.44 349</td>
<td>2197.87 349</td>
<td>2205.64 354</td>
<td>11.27 n.s.</td>
</tr>
<tr>
<td>Matrix reasoning</td>
<td>1842.22 347</td>
<td>1845.75 349</td>
<td>1845.93 350</td>
<td>1848.00 354</td>
<td>5.78 n.s.</td>
</tr>
<tr>
<td>Similarities</td>
<td>2150.00 347</td>
<td>2151.07 349</td>
<td>2151.21 350</td>
<td>2157.71 354</td>
<td>7.71 n.s.</td>
</tr>
<tr>
<td>Incomplete pictures</td>
<td>1681.34 347</td>
<td>1681.81 349</td>
<td>1681.85 350</td>
<td>1687.18 354</td>
<td>5.84 n.s.</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>1919.46 347</td>
<td>1920.33 349</td>
<td>1920.44 350</td>
<td>1930.52 354</td>
<td>11.06 n.s.</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>2675.27 347</td>
<td>2678.30 349</td>
<td>2678.60 350</td>
<td>2682.41 354</td>
<td>7.14 n.s.</td>
</tr>
<tr>
<td>Digit symbol coding</td>
<td>2964.08 342</td>
<td>2965.69 344</td>
<td>2965.99 345</td>
<td>2967.20 349</td>
<td>3.12 n.s.</td>
</tr>
<tr>
<td>Digit symbol free recall</td>
<td>1082.13 343</td>
<td>1082.29 345</td>
<td>1082.61 346</td>
<td>1092.84 350</td>
<td>10.71 n.s.</td>
</tr>
</tbody>
</table>

df = degrees of freedom; –2LL = twice the negative log likelihood; n.s. = not significant: when the increase in χ² is not significant, the most restrictive model is accepted; *an increase in χ² of more than 14.07 for Δdf = 7 is significant at the 0.05 level.

To test whether the above differences in mean scores indicated true differences, univariate analyses in Mx using twice the negative log-likelihood were run. The results for these analyses are presented in Table 4, from which it can be seen that comparison of model 4, the most parsimonious model, with model 1 did not cause a significant worsening of the fit for any of the WAIS-III subtests. In other words, for all subtests a model which estimates all means to be equal fits better than a model in which all means are estimated separately. There was no reason to believe that means of twins and singletons in our sample differed in IQ.

We did find, however, that comparison of model 4 (all means equal) with model 3 (separate means for twins and siblings) showed a significant worsening of the fit for subtests arithmetic and digit symbol-free recall, in the sense that on arithmetic singletons performed slightly better than both MZ and DZ twins, and on digit symbol-free recall MZ twins performed slightly better than both DZ twins and singletons. We also found that MZ twins performed significantly better than DZ twins on subtest letter–number sequencing.

Discussion

It has been suggested that twins have an intellectual disadvantage compared with singletons and that twin samples are not representative of the normal population. If true, this might influence generalisability of heritability estimates obtained in twin studies, for instance by a restriction of range of IQ scores. In the Record et al study a standard IQ score difference of 4.4 points was found between twins and singletons. Our study had enough statistical power to detect an effect of at least the same magnitude on each of the individual IQ subtests. We found, however, no evidence of a twin–singleton difference. In fact, means and standard deviations in our study showed no differences at all between twins and singletons. In the Record et al study, where these differences were found, a priori differences in social class or genetic background of twin families and singleton families could never be ruled out. Since our twins and singletons came from the same family, social class and genetic background were perfectly matched across twin families and singleton families.

Our results are in line with an earlier report by Kalman who administered the Wechsler Bellevue Scale to 134 twin pairs (aged 60–89 years), and compared the scores of these twins to standardised scores based on a comparable group of singletons. Kalman concluded that there was no significant difference between twins and singletons in measures of intellectual performance.

Although in our study no evidence was found for twin–singleton differences in intellectual ability, one cannot necessarily generalise from this in respect of personality, lifestyle, disease susceptibility or mortality rates. However, recent comparisons of twins and singletons on problem behaviour, mortality rates and psychiatric symptoms have not suggested twin–singleton differences in these fields either. All in all, significant disadvantages of twins in comparison with singletons seem to be implied rather than observed.

Acknowledgements

We gratefully acknowledge the financial support of the USF (grant number 96/22) and the HFSP (grant number rg0154/1998-B).
References


