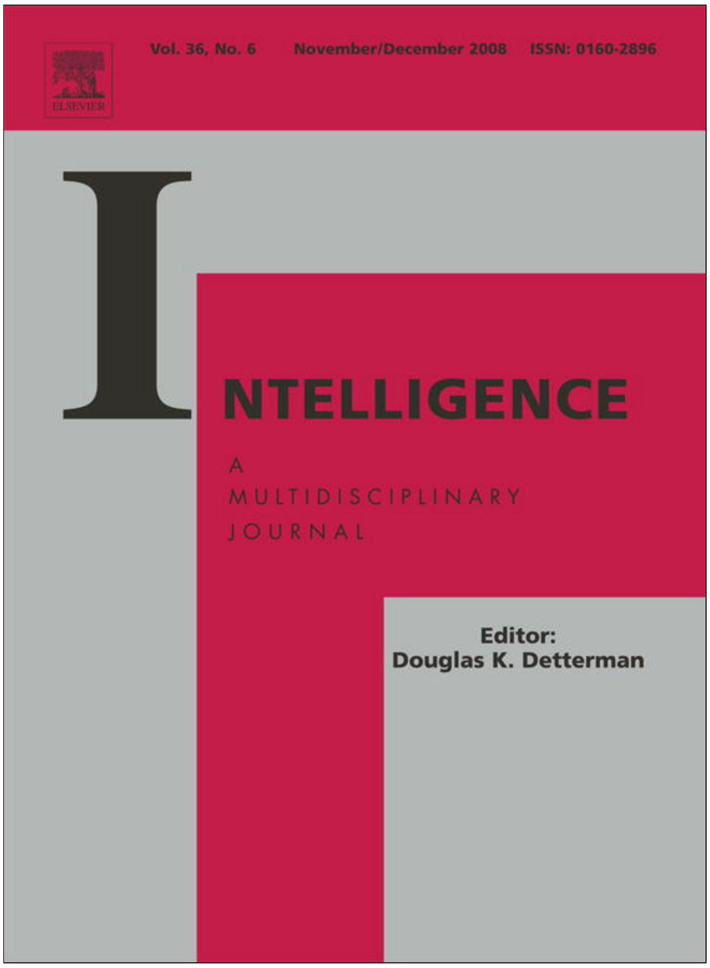
Vol. 36, No. 6
November/December 2008
ISSN: 0160-2896
INTELLIGENCE
A MULTIDISCIPLINARY JOURNAL
Editor:
Douglas K. Detterman

Available online at www.sciencedirect.com

ScienceDirect

Intelligence 36 (2008) 539–547

INTELLIGENCE

Do twins have lower cognitive ability than singletons?

Dinand Webbink [a,*], Danielle Posthuma [b], Dorret I. Boomsma [b], Eco J.C. de Geus [b], Peter M. Visscher [c]

[a] CPB Netherlands Bureau for Economic Policy Analysis, PO Box 80510, 2508 GM, The Hague, The Netherlands
[b] Department of Biological Psychology, Vrije Universiteit Amsterdam, The Netherlands
[c] Queensland Institute of Medical Research, Brisbane, Australia

Received 23 May 2007; received in revised form 7 December 2007; accepted 7 December 2007
Available online 1 February 2008

Abstract

Previous studies based on population cohorts born at least 35 years ago, have reported appreciable childhood cognitive deficits for twins. We compared longitudinal IQ scores from approximately 188,000 singletons and some 6000 twins who went to primary school in the Netherlands from 1994 to 2003. In addition, we used a family-based design in which IQ scores of *adult* twins ($N=196$) were compared with those of their adult singleton siblings ($N=589$). After correcting for such confounding factors as the year of testing, gender, age at the time of the test, and parents' education and ethnicity, twins aged 6 scored 16% of a standard deviation lower than non-twins in language and 17% of a standard deviation lower in arithmetic. For twins aged 8 the difference with non-twins in language and arithmetic reduced to 5% and 2% of a standard deviation and for twins aged 10 and 12 the differences were not statistically significant. For IQ scores, twins scored 0.09 points lower than non-twins at age 8 and 0.83 points lower at 10. However, twins scored higher at age 12 by 0.14 points. The only significant difference found was at age 10. Using the family-based adult sample, no differences in IQ scores were found between twins and their singleton siblings. These results suggest that in a very recent generation of school children in the Netherlands, there was a small but significant cognitive deficit for twins aged 6 and 8. However, the difference disappeared by the time the children were 12, and was also insignificant in the adult population. Previous studies, based on cohorts born more than 35 years ago in Britain, reported much larger cognitive deficits in twins. Whatever the reason of the cognitive deficit at age 6, which could include prenatal growth, shorter gestation and parental care, twins caught up and the cognitive cost of being a twin in the Netherlands seems to be minor and temporary.

Keywords: Cognitive ability; Twins; Singletons

1. Introduction

Cognitive ability is a key factor for important outcomes in life such as educational attainment, labour market opportunities and health in adulthood (Batty & Deary, 2004; Batty, Der, Macintyre, & Deary, 2006; Chandola, Deary, Blane, & Batty, 2006; Deary, Batty, & Gottfredson, 2005; Hart et al., 2003). Several studies, based on population cohorts born at least 35 years ago, found appreciable cognitive deficits for twins in childhood (Deary, Pattie, Wilson & Whalley, 2005; Ronalds, De Stavola, & Leon, 2005). This finding may have important public health and social implications (Batty et al., 2006; Hart et al., 2003; Starr et al., 2004; Stewart,

* Corresponding author. Tel.: +31 70 338 33 80.
E-mail address: webbink@cpb.nl (D. Webbink).

0160-2896/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2007.12.002

Deary, Fowkes, & Price, 2006; Taylor et al., 2005). However, a study based on recent birth cohorts found no difference in cognitive ability between twins and singletons (Christensen et al., 2006). An important question is whether these differences in cognitive ability no longer exist for recent cohorts and whether there are deficits of twins during childhood.

A general finding in most previous studies is that twins have lower cognitive ability than singletons (Deary, Pattie et al., 2005; Ronalds et al., 2005; Record, McKeown, & Edwards, 1970). The twins in a study of children who attended primary school in Aberdeen in 1962 had a cognitive deficit of more than 6 IQ points compared with singletons at ages 7 and 9 (Ronalds et al., 2005). The sample used in the analysis included 9832 singletons and 236 twins born between 1950 and 1956. The difference in mean IQ between twins and singletons within families was approximately 5 IQ points. Adjusting for the lower birth weight of twins halved the difference at age 7 and reduced it by 30% at age 9.

A study using the Scottish Mental Surveys of 1932 and 1947 reported a deficit of about 5 IQ points for twins at the age of 11 years (Deary, Pattie et al., 2005). To overcome the problem of selection bias due to a priori differences in social background between twins and singletons, this study examined two whole-population surveys, one of which also had information on social background. The total sample included approximately 2000 twins. A study of children born in Birmingham, United Kingdom between 1950 and 1954, based on 1082 twin pairs, found that twins had an average deficit in verbal reasoning scores at age 11 of 4.4 points (Record et al., 1970).

The findings from a study based on the Netherlands Twin Registry deviated from the general pattern. A comparison of 260 adult twins with 98 singleton siblings showed no significant difference in cognitive ability, even though a power analysis demonstrated that effects much smaller than those reported in previous studies could easily have been detected (Posthuma, De Geus, Bleichrodt, & Boomsma, 2000). A recent study of children born in Denmark during 1986–1988 also failed to find a difference between twins and singletons (Christensen et al., 2006). This study compared the school performance of 3411 twins with 7796 singletons between the age of 15 and 16.

A limitation of most studies is that they were based on populations of individuals born at least 35 years ago. The populations used in two recent studies (Deary, Pattie et al., 2005; Ronalds et al., 2005) were in fact born more than 50 years ago. Since then there has been considerable progress in the fields of obstetrics and neonatal care, which may have reduced these cognitive differences. It is also conceivable that the education system has evolved to better deal with children with cognitive deficits. However, Ronalds et al. (2005) argued that differences in cognition between twins and singletons are likely to remain due to the shorter gestation and impaired fetal growth that affect some twins. They stated that this “requires study of a more contemporary cohort”.

This is precisely the focus of our study. We used two data sets collected in the Netherlands. The first data set is a longitudinal data set of children in Dutch primary education born in the 1980s and 1990s. We were able to identify twins from large population based samples. Therefore, the number of twins in our study was much larger than in previous studies. We identified more than 6000 twins. In addition, the longitudinal character of our data enabled us to compare the cognitive development of twins with the development of singletons at the individual level. Hence, we could observe whether the differences found are permanent or change during the course of primary school.

We used a within school design to look for any difference in cognition between twins and singletons while controlling for various socioeconomic characteristics of the parents. In addition, we exploited the longitudinal character of the data to analyze cognitive development. We also investigated whether the group of twins observed in school at older ages is selective. If twins have lower cognitive ability, it is plausible that they will attend special education or repeat classes more often than singletons and this might bias the comparisons in later grades. The data came from representative samples of the Dutch population with an over-sampling of children from a lower socio-economic background.

Our second data set is a sample of Dutch twins and siblings from an ongoing study of cognition in adults (Posthuma et al., 2000). A total of 785 subjects from 316 families participated, of which 589 were twins, and 196 were siblings of a twin. All subjects were tested for IQ between 1996 and 2001. We used a within family design to investigate the differences in IQ between twins and singletons. The within family comparison has the added value that twins are compared with their singleton siblings. Singleton siblings make an ideal control group, because they are genetically similar and have a similar early family environment. Part of this sample (46%; 260 twins and 98 singleton siblings) was included in a previous paper (Posthuma et al., 2000), in which subtests of the WAIS-IIIR for the twins and singletons were compared. The full scale IQ, however, was not used, because Dutch standardization norms were unavailable at that time.

2. Methods

2.1. Subjects and data

Data were available from the longitudinal PRIMA survey in the Netherlands. This survey data is used to analyze the educational strategies and performance of the primary education system in the Netherlands (Driessen, Van Langen, & Vierke, 2004; Driessen, Van Langen, & Oudenhoven, 1994). We used the first five waves of the PRIMA survey including data on pupils, parents, teachers and schools. Each wave has approximately 60,000 pupils. The waves were from the years 1994, 1996, 1998, 2000, and 2002. The PRIMA project consists of a panel of approximately 600 schools, of which 180 schools are selected for the over-sampling of pupils with a lower social-economic background. Within each school, pupils in grades 2, 4, 6 and 8 (average age: 6, 8, 10, 12 years) were tested in language and arithmetic. Additionally, information on the social background was collected, and teachers were asked about the behaviour of the child in school. All participating pupils were assigned a personal ID which is used to link observations on the same pupil across waves. Approximately 50 to 60% of the pupils could be linked to a previous wave. Twins could be identified, because they share the same family name, date of birth, school and year of the survey (Webbink, Roeleveld, & Visscher, 2006).

The cognitive measures we used in the analysis were an IQ score and two achievement scores: one for arithmetic and one for language. All three tests were especially designed for the PRIMA survey.

The IQ test focused on non verbal intelligence and has two components: ‘composition of figures’ and ‘exclusion’. For the first component pupils had to compose a figure, for instance a square, from several irregular segments. For the exclusion test pupils had to choose one figure that did not fit into a sequence of figures. These two non verbal components have been chosen to measure intelligence unbiased by the socioeconomic background of the pupils (Driessen et al., 2004; Driessen et al., 1994). The IQ test was taken in grades 4, 6 and 8. For each grade there was a different version of the test. In the case of fourth grade, the composition of figures test had 17 items and the exclusion test had 20 items, for sixth and eighth grades there were 19 for composition and 15 for exclusion. The score on the test was the number of items correctly done. The same IQ test was used in all waves of the PRIMA survey. Due to the difference in difficulty and in the number of items in the tests, the tests were not comparable between grades. Recently, a more extended test for non-school related cognitive abilities, including the components ‘composition of figures’ and ‘exclusion’, has been favourably judged by the Committee On Test Affairs Netherlands (COTAN) of the Dutch professional association of psychologists (NIP) (Van Batenburg & Van der Werf, 2004). The tests for languages and arithmetic were developed by the CITO group (Kamphuis, Mulder, Vierke, Overmaat, & Koopman, 1998). The language test for children in second grade, which is equivalent to infant school, measures the understanding of words and concepts. The arithmetic test for these children focused on the sorting of objects. These tests could be taken in class. The test for children in grades 4, 6 and 8 all came from a system for following pupil achievements in primary education developed by the CITO group. The tests for the same grade levels were identical each year. This ensured that the comparison of achievement levels over time was possible. The scores were also comparable between grades.

The scales of the raw scores for language and arithmetic have no clear meaning. We have therefore opted to transform these scores for each test into wave specific standardized scores, having mean zero and standard deviation one. IQ scores were normalised to a mean of 100, with a standard deviation of 15.

Since the start of the project, several tests have changed. In the PRIMA project these scores have been calibrated to be on the same scales as the previous test versions. This ensures that they can still be compared to the results from earlier years. It should be noted that the comparability over time is hampered by other differences between waves. In the first wave, tests were taken early in the school year. In the second wave, tests were taken halfway through the school year. In the first two waves, tests were administered by an external examiner, while in the third wave the class teacher administered the tests. Because these differences might affect our findings we control for the year of the survey in all regressions.

At the individual level we controlled for gender, age (in survey year), education levels, and country of birth of the mother and father, and the pupil’s so-called weight factor assigned by the funding scheme for primary schools. The Dutch funding scheme for primary schools distinguishes several groups of disadvantaged pupils. The most important groups are Dutch pupils with lower educated parents and pupils with an ethnic minority background. Pupils not belonging to a disadvantaged group enter the funding scheme with a weight factor equal to unity. Dutch pupils of poorly educated parents have a weight equal to 1.25 and pupils from an ethnic minority have a weight factor of 1.9. Schools receive 25% additional funding for pupils with a weight of 1.25, and 90% additional funding for these pupils with a weight of 1.90.

The first five waves of the PRIMA project contained information on 290,910 pupils of which 6029 were identified as twins. Because of the longitudinal character of the project, the number of unique pupils was smaller. The first five waves contain 192,102 unique individuals of which 3894 were twins. In the identification procedure for the twins we also made use of the longitudinal character of the data. This explains the odd total number of twins identified.

Data on the language and arithmetic tests were available for approximately 270,000 pupils. The IQ test was not taken in second grade, and therefore data are available for 191,357 pupils. The data on gender was missing for 12,820 observations. The weight factor for the financing of schools was measured in five categories and is missing for 14,157 pupils. Education of the father and mother was measured in four categories and was missing for 49,685 fathers and 33,596 mothers. The parents' country of birth was sorted into 14 different categories and was missing for 27,326 fathers and 33,596 mothers. These explanatory variables were included as categorical variables in the analysis. For the missing values of a certain variable we used an additional category. The date of birth was missing for 5767 pupils.

In the case of our second data set of adult Dutch twins and siblings, subjects were recruited from the Netherlands Twin Register to participate in the cognition study and gave written informed consent. A total of 785 subjects (342 males) from 316 families participated, of which 589 were twins, and 196 were siblings of a twin. Family-size ranged from 1 to 8, with a mean of 2.51 (SD .84) offspring per family. The mean age of the twins was 37.6 years (SD 12.9), and of the siblings 37.8 years (SD 13.1).

The institutional review board of the Vrije Universiteit Medical Centre approved DNA sampling for zygosity determination and cognitive testing. The Dutch adaptation of the WAISIII-R (Wechsler, 1997) was used to assess IQ and consisted of eleven subtests, nine of which can be used to calculate Full Scale IQ score according to the WAIS-III manual: subtests information, similarities, vocabulary, arithmetic, letter–number sequencing, picture completion, block design, matrix reasoning and digit-symbol substitution.

The current sample is larger than the sample used for standardization norms in the WAIS-III manual. We found small significant effects of age and sex on the scaled scores, and therefore conducted all analyses on the residual FSIQ score (i.e., corrected for age and sex).

2.2. Statistical methods

We compared the mean cognitive values for the IQ, language and arithmetic tests measured in grades 2, 4, 6 and 8, and the values of the control variables of singletons and twins overall. A major concern when comparing twins with singletons is that there may be differences in the characteristics of the family and socioeconomic environment that might also be related to cognitive ability. In other studies, the effects of these potential confounders were eliminated by using a within family design (Ronalds et al., 2005; Posthuma et al., 2000).

The PRIMA data did not contain information on siblings, but we nonetheless were able to use a similar approach. Fixed effects linear regressions were used to estimate the mean difference between the cognitive test score of twins and singletons in each school. These estimates control for fixed school characteristics and the standard errors are adjusted for clustering of pupils within schools.

We also note that the Dutch primary education system allows parents to choose which school their children attend. This has the effect of sorting pupils with comparable socioeconomic characteristics into the same schools. Therefore, it is to be expected that the school fixed effect will also capture family effects to some extent. In our analysis we also adjusted for observed potential confounders by introducing them into the models.

In our analysis we were also able to exploit the longitudinal aspect of the data. First we investigated whether the comparison between twins and singletons in later grades is biased. We estimated probit models of participation in the next wave of the survey to check if twins are less likely to participate in the next wave. We also compared the cognitive development of those twins and singletons for which we have cognitive measures from several waves of the survey.

Our analysis of the cross-sectional adult sample from the NTR is based on a within family design. This design tests for differences between the group mean and variance for twins and the group mean and variance for singletons, while correcting for dependency in the data. Analyses were carried out using the structural equation modelling implemented in the statistical software package Mx (Neale et al., 2003). Mx optimizes the raw maximum likelihood function for a given model. Nested models can be formally compared using the likelihood-ratio test statistic.

3. Results

Table 1 shows separate sample statistics for twins and for singletons of the dependent and explanatory variables in all four grades and in the adult sample. Twins had lower scores in languages and arithmetic in second grade. In fourth grade twins only scored lower in

Table 1
Sample means (standard deviations) for dependent and explanatory variables in grades 2, 4, 6 and 8 (longitudinal data) and in adulthood, separately for singletons and twins

	Grade 2		Grade 4		Grade 6		Grade 8		Adult	
	Singleton	Twin	Singleton	Twin	Singleton	Twin	Singleton	Twin	Singleton	Twin
Language	−1.17	−1.22	−0.18	−0.15	0.41	0.43	1.05	1.10	n.a.	n.a.
	(0.57)	(0.56)	(0.60)	(0.60)	(0.55)	(0.54)	(0.59)	(0.59)		
Arithmetic	−1.15	−1.19	−0.37	−0.39	0.50	0.49	1.10	1.24	n.a.	n.a.
	(0.54)	(0.53)	(0.54)	(0.54)	(0.37)	(0.37)	(0.59)	(0.37)		
IQ	n.a.	n.a.	100.0	100.7	100.0	100.0	100.0	100.7	103.9	103.5
			(15.0)	(14.9)	(15.0)	(14.8)	(15.0)	(14.4)	(11.7)	(11.4)
% Girl	48.5	50.8	49.6	50.1	49.9	52.0	50.2	54.0	53.6	57.4
Age at test	5.83	5.84	7.95	7.95	9.99	9.98	11.99	11.97	37.75	37.56
(years)	(0.42)	(0.39)	(0.50)	(0.44)	(0.53)	(0.48)	(0.53)	(0.48)	(13.06)	(12.93)
Weight factor (%)										
1.0%	48.2	54.3	46.3	53.9	45.1	52.2	44.3	49.6	n.a.	n.a.
1.25%	20.2	19.7	22.4	20.6	24.5	22.1	26.8	25.7	n.a.	n.a.
1.9%	26.2	20.7	25.8	20.6	24.7	20.4	23.5	20.7	n.a.	n.a.
% foreign country of birth										
Father	36.7	31.3	35.7	29.5	35.0	29.4	33.8	29.1	n.a.	n.a.
Mother	33.4	27.4	32.5	27.3	31.9	26.8	30.8	25.9	n.a.	n.a.
Parents education										
Father (1–4)	2.48	2.58	2.46	2.59	2.45	2.57	2.43	2.56	n.a.	n.a.
Mother (1–4)	2.40	2.49	2.38	2.50	2.36	2.45	2.33	2.41	n.a.	n.a.
Observations	75,056	1791	73,989	1678	69,432	1410	66,404	1150	196	589

languages. In the later grades we observe that twin scores were either nearly equal to or are better than singletons. In the adult sample the IQ score of singletons was 0.5 point higher.

The proportion of girls was higher among twins in the later grades. Additionally, twins tended to have more highly educated parents, who were more often born in the Netherlands. The weight factor for the financing of primary schools, which is based on the education and ethnicity of the parents, confirmed that twins fall into category 1.0 (regular Dutch parents) more often, and fall less into category 1.25 (poorly educated Dutch parents) or category 1.9 (poorly educated foreign parents). These differences were found in all grades. Hence, twins had a higher socioeconomic background than singletons. It should be noted that approximately one third of the parents were born in a foreign country. This high share reflected the over-sampling of 'disadvantaged pupils' in the PRIMA project from which the data were taken.

Table 2 shows the difference in test scores between twins and singletons from the fixed effects regression adjusted for various potential confounders. We first estimated a model that only controls for the year of the survey and added potential confounders in subsequent models. For the IQ test we found very small differences, which were only significant in sixth grade. It should be noted that the largest difference we found is less than one IQ-point. This is smaller than the 6 IQ points difference Ronalds et al. (2005) reported.

For the language and arithmetic tests we found that twins score significantly lower in second grade. The differences in the full model were approximately 9% of a standard deviation of the distribution for all grades. Again, these are small differences compared to previous findings. In later grades the cognitive deficit of twins became smaller and turned insignificant. In eighth grade we even observed higher test scores for twins than for singletons. These estimates suggest a catch up in cognitive ability of twins. The estimates were quite robust to the inclusion in the model of additional controls.

In general, the relative score of twins decreased as we included additional controls in our model. However, the changes in the estimates were very small even for variables which are known to be very important predictors of educational performance, such as parents' education. This could be explained by the fact that in the model, which only controls for the year of the survey, we already adjusted for a school fixed effect. This picked up a lot of variation in the socioeconomic background of the pupils.

We note that the survey design involved the sampling of pupils from grades and not from cohorts. This is unfortunate because repeating a grade is a fairly

Table 2
Estimates of the difference (standard errors) in cognitive ability between singletons and twins in grades 2, 4, 6 and 8 after fitting confounding factors

	Grade 2	Grade 4	Grade 6	Grade 8
IQ	*			
Adjusted for				
Year		0.331 (0.377)	−0.536 (0.415)	0.428 (0.455)
+Gender		0.330 (0.376)	−0.566 (0.415)	0.423 (0.455)
+Age		0.292 (0.378)	−0.583 (0.415)	0.366 (0.452)
+Weight factor		0.007 (0.374)	−0.737 (0.412)	0.344 (0.449)
+Parents education		−0.073 (0.372)	−0.773 (0.411)	0.183 (0.446)
+Parents ethnicity		−0.086 (0.371)	−0.827 (0.410)	0.143 (0.446)
+Observations		65,975	62,147	59,774
Language				
Year	−0.076 (0.013)	−0.002 (0.014)	−0.002 (0.014)	0.028 (0.017)
+Gender	−0.078 (0.013)	−0.003 (0.014)	−0.004 (0.014)	0.028 (0.017)
+Age	−0.082 (0.013)	−0.004 (0.014)	−0.007 (0.014)	0.025 (0.016)
+Weight factor	−0.087 (0.012)	−0.021 (0.013)	−0.014 (0.013)	0.023 (0.016)
+Parents education	−0.087 (0.012)	−0.025 (0.013)	−0.017 (0.013)	0.016 (0.016)
+Parents ethnicity	−0.089 (0.012)	−0.028 (0.013)	−0.017 (0.013)	0.014 (0.016)
Observations	68,948	69,881	65,894	63,266
Arithmetic				
Year	−0.077 (0.012)	0.003 (0.012)	−0.005 (0.010)	0.019 (0.011)
+Gender	−0.078 (0.012)	0.003 (0.012)	−0.003 (0.010)	0.024 (0.011)
+Age	−0.083 (0.012)	0.002 (0.012)	−0.005 (0.009)	0.022 (0.010)
+Weight factor	−0.086 (0.011)	−0.007 (0.012)	−0.009 (0.009)	0.020 (0.010)
+Parents education	−0.088 (0.011)	−0.011 (0.012)	−0.010 (0.009)	0.015 (0.010)
+Parents ethnicity	−0.090 (0.011)	−0.012 (0.012)	−0.011 (0.009)	0.014 (0.010)
Observations	69,559	69,738	64,256	61,169

*The IQ test was not taken in grade 2.

common phenomenon in Dutch primary schools. This has the potential to introduce selection bias into our analysis. Only in the case of the transition from first to second grade is repeating a year not possible. Therefore, the findings after second grade could be biased by panel attrition if twins are more likely to leave schools or repeat a grade than singletons.

A rough indicator for selective attrition is the sample mean on age. Table 1 shows that twins did not differ from singletons in age. We estimated probit models on the participation in the next wave to investigate the selective attrition of twins. The dependent variable in these models was participation in the next wave. As explanatory variables we used all controls from Table 2 and the twin dummy. In a second specification we added the test scores in the previous waves. We found no evidence for selective attrition of twins from second to fourth grade or from fourth to sixth grade. We conclude that the finding in Table 2 that twins catch up with their singleton peers, which primarily occurs from second to fourth grade, could not be explained by bias due to differences in school careers of twins and singletons. In the case of the transition from sixth to eighth grade, we found that twins were approximately 3% points more likely to drop out of the sample.

We also investigated whether the twins' mean cognitive deficit in second grade changed as their primary education progresses. We did this using sub samples of pupils who have test scores available in multiple waves. More specifically, we tracked pupils' performance on the language and arithmetic tests over several grades. In the analysis we used the same explanatory variables as those used in Table 2 and, as in Table 2, we also included a fixed school effect. The second column in Table 3 shows the results for the sample of pupils with test scores observed in the second and fourth grades. The third column shows the results for pupils with test scores observed in grades 2, 4 and 6 and the fourth column for pupils with test scores observed in grades 2, 4, 6 and 8. These estimates demonstrate how twins caught up with their singleton peers. For the language test we found that the cognitive deficit of twins gradually changed into a cognitive surplus for twins during the course of their primary education. In each subsequent grade the relative position of twins improved. For arithmetic we also found that twins caught up, although the pattern was less clear than for the language test.

Table 3
Estimates (SE) of cognitive improvement of twins in grade 2, 4, 6 and 8 using longitudinal samples

	Sample observed in		
	Grades 2 and 4	Grades 2, 4 and 6	Grades 2 to 8
*Language**			
Grade 2	−0.060 (0.019)	−0.065 (0.029)	−0.039 (0.046)
Grade 4	−0.035 (0.021)	−0.001 (0.031)	0.022 (0.052)
Grade 6		0.013 (0.030)	0.053 (0.051)
Grade 8			0.091 (0.054)
Highest–lowest grade	0.026 (0.024)	0.078 (0.035)	0.130 (0.060)
Observations	27,530	11,576	4684
*Arithmetic**			
Grade 2	−0.084 (0.017)	−0.057 (0.027)	−0.055 (0.042)
Grade 4	−0.016 (0.017)	0.007 (0.026)	0.031 (0.045)
Grade 6		−0.034 (0.021)	−0.001 (0.034)
Grade 8			0.008 (0.031)
Highest–lowest grade	0.068 (0.018)	0.023 (0.026)	0.063 (0.042)
Observations	27,641	11,134	4155

*Estimates shown are the coefficient of a twin dummy in regression models with a school fixed effect and all controls from Table 2 were used.

In the family-based adult samples a mean IQ of 103.30 (SD 11.33) was estimated for twins and 104.57 (11.85) for singletons. Neither the variance nor the means were significantly different for twins and singletons (χ^2 (1)=0.63; p=0.43 and χ^2 (1)=2.26; p=0.13, respectively).

4. Discussion

A small but significant difference in cognitive ability was found between twins and singletons for 6 year olds from a recent generation of school children in the Netherlands. This difference decreased and was statistically insignificant in later grades. This difference found at age 6 is, however, less than one IQ-point. For both tests in language and arithmetic, it was found that twins score 9% of a standard deviation lower than singletons at age 6, where the standard deviation has been normalized based on the distribution of test scores in all grades. When measured as a percentage decline in terms of the standard deviation normalized based only on the sample of six year olds, we obtained a decline of 16% for languages and 17% for arithmetic. Twins caught up in later grades and this cannot be explained by sample attrition due to differences in school careers between twins and singletons. Moreover, in the longitudinal analysis of the sample of twins and singletons who are observed in at least two waves, we also found that twins caught up with singletons in language and arithmetic. That twins do eventually catch up was confirmed in the family-based adult sample, where no difference was found between twins and their singleton siblings.

Most previous studies have found appreciable differences between twins and singletons. For instance, a cognitive deficit of twins of more than 6 IQ points at ages 7 and 9 was reported recently (Ronalds et al., 2005), which is 35 to 40% of a standard deviation. Our largest estimate of cognitive deficit was 0.827 IQ points at age 10 (Table 2), or roughly 6% of a standard deviation. A remarkable finding in the Ronalds study was that the sample means of the singletons on the age adjusted IQ scores were much higher than the expected sample mean of 100 (108 for age 7 and 112 for age 9). This raises the question of whether the apparent exclusion of approximately 2000 children from their sample may have caused a bias in their estimates.

Our findings are consistent with a recent study of twins and singletons aged 15 or 16 in Denmark (Christensen et al., 2006). Although speculative, one explanation for not finding a difference in cognitive ability between singletons and twins in the Dutch population might be found in the current Dutch health and education systems. It may well be the case that the present Dutch health and education systems are more efficient in reducing differences between individuals in health and cognition status that occur early in life than their British counterparts were 35 to 80 years ago. However, we are not aware of specific school policies that take a possible early cognitive deficit of twins into account.

Some cautionary notes about this study may be in order. In our study we identified twins based on a matching strategy. This strategy might produce false twin pairs if pupils in the same school and grade with the same family name and date of birth are not related. In Webbink et al. (2006) the authors provided evidence that the accuracy of the matching procedure is high. Hence, we expect the bias from the measurement error in identifying twins to be small. Another potential pitfall arises if the proportion of twins in special education exceeded the proportion of singletons in special education. In this case, we probably underestimated the difference in cognitive abilities, because our study did not include pupils in special education. This means that our matching strategy would not identify twin pairs if one of the twins is in special education. Data from the Dutch twin registry (NTR) show that 3.2% of twins were in special education at age 7, 5.9% at age 10 and 5.6% at age 12. The figures for the total population were slightly higher, for instance in 1994 to 1995 3.8% of pupils were in special education at age 7, 7.4% at age 10 and 7.7% at age 12 (Statistics Netherlands). These findings are quite similar to the Danish study where the proportion of missing test scores was equal for twins and singletons (Christensen et al., 2006).

Our approach also suffers from the drawback that we cannot identify twins if they attend different schools. Data from the Dutch twin registry show that at age 5 only 2.3% of twins go to different schools and at age 12 5.7% go to different schools (Van Leeuwen, Van den Berg, Van Beijsterveldt, & Boomsma, 2005). This suggests that the possible bias from unobserved separated twins will be small. For the identification of twins in later grades we also used the longitudinal character of the data by using the personal ID. The samples of pupils who were observed in multiple grades are smaller, but the estimates of the cognitive deficits were quite similar to the estimates for the total sample. In any case, even if the real proportion of twins in special education were double the proportion reported by the Dutch Twin Registry, we would not find the reported 5 IQ points deficit for twins in the literature. If 10%, and not 5%, of twins were in special education and if their average IQ = 70, then the mean for all twins would be $0.10*70+0.90*100=97$. This would still only represent a cognitive difference of 3 IQ points. If anything, the available figures on special education suggest that there are proportionally more non-twins in special education. Hence the mean of the singletons is biased by the same amount or by even more than that of the twins. We therefore do not expect that adjusting for special education would change our conclusions.

One possible concern about our findings based on the adult sample is the potential for recruitment bias in twin studies, as suggested by Ronalds et al. (2005). However, when twins and singletons are compared from the same family as we have done here, this ensures ideal matching on genetic and social background. Because IQ scores cluster in families, recruitment bias is expected to affect both twins and their singleton siblings in a similar manner and is therefore unlikely to explain a lack of twin–singleton differences in IQ.

Finally, we note that our findings on pupils in primary education may be biased by potential confounders such as unobserved postnatal family and socioeconomic characteristics, because we could not use a within family design as has been used by others (Ronalds et al., 2005; Posthuma et al., 2000). However, in our fixed effects models we controlled for parents' education and country of birth. Since the publication of the Coleman report, it is widely accepted that these variables are important predictors of cognitive ability and educational performance (Coleman, 1966). Our estimates show that adjusting for these key variables only led to small changes in the estimated difference between twins and singletons. If the change in the estimates due to the adjustment for the observed confounders can be seen as a guide for the change in the effects due to unobserved confounders, it seems implausible that there are appreciable cognitive differences between twins and singletons for the cohorts that we study (Altonji, Elder, & Taber, 2005). Moreover, previous research showed that the estimates changed only slightly after taking account of fixed family effects (Ronalds et al., 2005). In the adult sample, we were in fact able to control family and socioeconomic characteristics, because this is a family-based sample in which twins were compared with their singleton siblings. Findings on the adult sample corroborated our findings on pupils in primary education.

Competing interests

PMV declares that he is the proud father of school-aged twin boys.

Acknowledgements

Part of PMV's research is funded by Australian National Health and Medical Research Council grants (389892 and 442915). DP is supported by NWO/MaGW Vernieuwingsimpuls 016-065-318. The funding sources had no involvement in this study. We thank Rob Luginbuhl for many helpful editorial changes.

References

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*, 151−184.

Batty, G. D., & Deary, I. J. (2004). Early life intelligence and adult health. *British Medical Journal*, *329*, 585−586.

Batty, G. D., Der, G., Macintyre, S., & Deary, I. J. (2006). Does IQ explain socioeconomic inequalities in health? Evidence from a population based cohort study in the west of Scotland. *British Medical Journal*, *332*, 580−584.

Chandola, T., Deary, I. J., Blane, D., & Batty, G. D. (2006). Childhood IQ in relation to obesity and weight gain in adult life: the National Child Development (1958) Study. *International Journal of Obesity (London)*.

Christensen, K., Petersen, I., Skytthe, A., Herskind, A. M., McGue, M., & Bingley, P. (2006). Comparison of academic performance of twins and singletons in adolescence: Follow-up study. *British Medical Journal*, *333*, 1095.

Coleman, J. S. (1966). *Equality of educational opportunity.* Government Printing Office.

Deary, I. J., Batty, D., & Gottfredson, L. S. (2005). Human hierarchies, health, and IQ.*Science*, *309*, 703 (author reply 703).

Deary, I. J., Pattie, A., Wilson, V., & Whalley, L. J. (2005). The cognitive cost of being a twin: two whole-population surveys. *Twin Research Human Genetics*, *8*, 376−383.

Driessen, G., Van Langen, A., & Oudenhoven, X. (1994). *De toetsen voor de cohort Primair onderwijs, verantwoording.* Nijmegen ITS.

Driessen, G., Van Langen, A., & Vierke, H. (2004). *Basisrapportage PRIMA-cohortonderzoek, Vijfde meting 2002–2003. (Report on PRIMA-longitudinal research project, Survey 2002–2003).* Nijmegen.

Hart, C. L., Taylor, M. D., Davey Smith, G., Whalley, L. J., Starr, J. M., Hole, D. J., et al. (2003). Childhood IQ, social class, deprivation, and their relationships with mortality and morbidity risk in later life: Prospective observational study linking the Scottish Mental Survey 1932 and the Midspan studies. *Psychosomatic Medicine*, *65*, 877−883.

Kamphuis, F., Mulder, L., Vierke, H., Overmaat, M., & Koopman, P. (1998). *De relatie tussen PRIMA toetsen en toetsen uit het CITO-leerling volgsysteem.* Nijmegen ITS.

Neale, M. C., Boker, S. M., Xie, G., et al. (2003). *Mx: Statistical modelling* (6th ed.). Richmond, Va: VCU Department of Psychiatry.

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

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.

Ronalds, G. A., De Stavola, B. L., & Leon, D. A. (2005). The cognitive cost of being a twin: Evidence from comparisons within families in the Aberdeen children of the 1950s cohort study. *British Medical Journal*, *331*, 1306.

Starr, J. M., McGurn, B., Whiteman, M., Pattie, A., Whalley, L. J., & Deary, I. J. (2004). Life long changes in cognitive ability are associated with prescribed medications in old age. *International Journal of Geriatric Psychiatry*, *19*, 327−332.

Stewart, M. C., Deary, I. J., Fowkes, F. G., & Price, J. F. (2006). Relationship between lifetime smoking, smoking status at older age and human cognitive function. *Neuroepidemiology*, *26*, 83−92.

Taylor, M. D., Hart, C. L., Smith, G. D., Starr, J. M., Hole, D. J., Whalley, L. J., et al. (2005). Childhood IQ and social factors on smoking behaviour, lung function and smoking-related outcomes in adulthood: Linking the Scottish Mental Survey 1932 and the Midspan studies. *British Journal of Health Psychology*, *10*, 399−410.

Van Batenburg, T., & Van der Werf, M. (2004). *NSCCT. Niet Schoolse Cognitieve Capaciteiten Test, Voor groep 4, 6 en 8 van het basisonderwijs. Verantwoording, normering en handleiding.* Groningen GION.

Van Leeuwen, M., Van den Berg, S. M., Van Beijsterveldt, C. E. M., & Boomsma, D. I. (2005). Effects of twin separation in primary school. *Twin Research and Human Genetics*, *8*, 384−391.

Webbink, D., Roeleveld, J., & Visscher, P. M. (2006). Identification of twin pairs from large population based samples. *Twin Research and Human Genetics*, *9*, 496−500.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale-III.* San Antonio, TX The Psychological Corporation.