Achievement test performance in children conceived by IVF

L. Mains¹, M. Zimmerman², J. Blaine¹, B. Stegmann¹, A. Sparks¹, T. Ansley³, and B. Van Voorhis¹,*

¹Department of Obstetrics and Gynecology, University of Iowa Hospitals and Clinics (UIHC), 200 Hawkins Drive, Iowa City, IA 52242, USA
²Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA 52242, USA
³Iowa Testing Programs, University of Iowa, Iowa City, IA 52242, USA

*Corresponding address. Tel: +1-319-356-4536; E-mail: brad-van-voorhis@uiowa.edu

Submitted on May 4, 2010; resubmitted on June 23, 2010; accepted on July 5, 2010

BACKGROUND: Long-term follow-up studies of children conceived by IVF are limited. We examine academic performance on standardized tests [Iowa Tests of Basic Skills/Educational Development (ITBS/ITED)] of children conceived by IVF.

METHODS: Parents of children 8–17 years of age at the onset of the study (March 2008) who were conceived by IVF at the University of Iowa Hospitals and Clinics and living in the state of Iowa were contacted by mail. Parents completed questionnaires on their child’s health and education and parental education. ITBS/ITED scores from school grades 3–12 were obtained on IVF children and a group of anonymous children matched by grade, year, gender and school district. Scores were analyzed using linear mixed models.

RESULTS: Four hundred and ninety-seven couples were contacted. Two hundred and ninety-five couples (463 children) agreed to participate (59.4% of parents), with ITBS/ITED scores available on 423 children (91.4% of participants). IVF children scored higher than the national mean (P, 0.0001) across all grades and subtests and higher than their matched peers for grades 3–11. A trend toward lower test scores in multiple gestations was present (but not significant). Factors found to affect test scores included parental level of education, maternal age, divorce and child’s BMI. Cryopreservation, length of embryo culture and method of insemination did not affect scores.

CONCLUSIONS: IVF children scored higher on standardized tests than their matched peers, suggesting that IVF does not have a negative effect on cognitive development. However, long-term follow-up of IVF children is still limited. Further research should be performed on the effect of multiple gestation on academic performance.

Key words: IVF / achievement / cognitive development / multiple gestation / academic performance

Introduction

Since the inception of IVF in 1978, there has been a remarkable increase in the number of IVF cycles worldwide. Today 1% of babies born in the USA, nearly 2% in the UK and almost 4% in Denmark/Finland are conceived through this technology. Although IVF is generally considered safe, numerous studies have identified several short-term risks to children conceived by IVF. Adverse perinatal outcomes and certain birth defects are among these risks, even when comparing singleton babies (Helmerhorst et al., 2004; Jackson et al., 2004; Olson et al., 2005).

The cause of these adverse outcomes remains poorly understood. Studies on IVF children may be affected by ascertainment or treatment biases resulting from more careful scrutiny of babies conceived by IVF. The increased prevalence of adverse outcomes may also be attributable to some factor related to the parents’ diagnosis of infertility (Zhu et al., 2006; Romundstad et al., 2008). However, recent studies demonstrating changes in gene imprinting from in vitro culture of embryos raise concern about the effect of the IVF process itself on these children (Manipalviratn et al., 2009). The full implications of imprinting defects are currently not known, but there is evidence of a link between several imprinted genes and cognitive and neurologic development (Skuse et al., 1997; Isles and Wilkinson, 2000).

The known short-term risks in IVF children and recent evidence of gene imprinting defects justify longer follow-up studies, although currently such studies are limited, particularly beyond 8 years. In a review of controlled studies, Ludwig et al. (2006) concluded that much of the data in this area are conflicting, and data that do suggest a difference are often confounded by multiple gestation and associated prematurity. However, there appears to be evidence of an increased risk of neurologic problems in IVF children, specifically a small increase in cerebral palsy even when limited to singletons (Ericson et al., 2002; Stromberg et al., 2002; Lidgard et al., 2005). Fewer studies have examined more subtle measures of cognitive
development, and those that have are limited by small sample size, short length of follow-up, invalidated measures of cognitive ability or lack of a suitable control group (Cederblad et al., 1996; Olivennes et al., 1996, 1997; Levy-Shiff et al., 1998; Place and Englert, 2003; Bonduelle et al., 2005; Wagenaar et al., 2008a,b; Wennerholm et al., 2009). Limited long-term follow-up creates a challenge for physicians when counseling patients about possible adverse outcomes associated with IVF.

The goal of this study was to examine cognitive function in children conceived by IVF using the results of standardized tests through the Iowa Tests of Basic Skills (ITBS, grades 3–8) and the Iowa Tests of Educational Development (ITED, grades 9–12). ITBS and ITEd are achievement tests taken by virtually all children in the state of Iowa from grades 3 through 12 as well as widely used throughout the USA. The tests are recognized nationally as valid and reliable measures of academic achievement and even predictors of adult IQ (Spinks et al., 2007). Results on these tests have been found to correlate with school grades, college grade point average and college admission test scores including the Scholastic Aptitude Test, the American College Test and the Cognitive Abilities Test, among others (Lohman and Hagen, 2001). The results of this study would allow us to better inform parents who are considering IVF of the long-term implications related to cognitive development.

**Materials and Methods**

This was a retrospective cohort study approved by the University of Iowa Institutional Review Board. The IVF database at the University of Iowa Hospitals and Clinics was screened for IVF cycles with a cycle start date between 1 January 1989 and 31 December 1998 to search for children who were currently (March 2008) 8–17 years of age. Only test scores from the state of Iowa were available to us, and written parental consent was required to obtain a child’s test score. Therefore, cycles were eliminated if patients did not live in Iowa, their address could not be obtained, they had requested no contact or had a known stillbirth or neonatal death.

Mothers were mailed questionnaires about their children’s educational development as well as parental level of education. Specifically, parents were asked if their child had skipped or been held back a grade, had received extra help or advanced education or had been given a diagnosis of attention deficit or autism. The highest level of completed education [high school/general educational development (GED), college or graduate school] was requested for each parent.

Non-responders from the initial mailing were contacted via a second mailing and subsequently via telephone to increase participation. ITBS/ITED scores of enrolled children were requested for grades 3–12. Scores which were provided included National Standard Score (NSS) and National Percentile Rank (NPR) for Reading Total (RT), Language Total, (LT), Math Total (MT) and a composite of these three, Core Total (CT). Scores were also obtained on Reading Vocabulary (RV), one component of RT that has been found to most closely correlate with traditional IQ (Forsyth et al., 2003). A control score for each IVF child was generated in the following manner: a cohort of children matched to each IVF child by gender, grade level, calendar year the test was taken and school district was identified; test scores for these children were obtained for the five tests cited above; the means of these scores were calculated and used as the IVF child’s control scores. When fewer than 20 children were available in a control group, mean control scores were not generated to protect the privacy of students in that group.

We calculated that to detect a difference of five points in ITBS scores, we needed about 100 participants in each group for 80% power and a 95% confidence interval. We anticipated that we would safely meet this requirement as 50% participation would result in nearly 400 children in our IVF group.

**Statistical analysis**

Linear mixed model analysis for repeated measures was used to estimate the mean NPR at each grade and test the null hypothesis of Ho: mean NPR = 50. By performing this test, we were comparing the ITBS/ITED score of IVF children to the national norm. The linear mixed model included grade as a fixed effect to test if the mean difference from control changed with grade. Mother identification and cycle number were included in the model as random effects with the covariance structure fitted for ITBS/ITED scores over grade levels from the same child within mother and cycle to account for the correlation of scores from the same child.

To compare the scores of IVF children to their matched controls, a normalized score (z-score) was computed which was the difference between the child’s score and the average control score divided by the SD of the matched control scores. The NSS was used for the calculation of these z-scores. Similar to the analysis for NPR, linear mixed model analysis for repeated measures was used to estimate the mean normalized score (z-score) at each grade and test the null hypothesis of Ho: mean z-score = 0. By performing this test, we were comparing the ITBS/ITED score of IVF children to their gender–grade–year–district-matched control group.

The linear mixed model was also used to test the effect of the various parent, child and IVF cycle factors on the z-score. z-scores from all grades were included in this analysis. The fixed effects in the linear mixed model were the parent, child or IVF cycle factor of interest and grade. Parental factors analyzed included: maternal age at the birth, mother’s level of education, father’s level of education and marital status; child factors included: gestational age at birth, birthweight, type of delivery (vaginal, C-section, gender, child’s current BMI, gestation type (singleton, twins, triplets, quadruplets), diagnosis of attention-deficit hyperactivity disorder (ADHD) or autism, use of extra help services and use of the talented and gifted program; IVF cycle factors included: cycle type (fresh embryo transfer versus frozen embryo transfer), insemination type (standard versus ICSI) and length of embryo culture [0–1 days (gamete intra-Fallopian transfer or zygote intra-Fallopian transfer, ZIFT) versus 2–3 days (IVF)].

Statistical significance was set at $P < 0.05$.

**Results**

Screening the IVF database and eliminating cycles for the reasons listed in the Materials and methods section resulted in 794 children (564 IVF cycles, 497 women). Our initial response rate of 40.0% was increased to 51.0% through a second mailing and then to 59.4% through telephone contact resulting in 463 children (335 cycles, 295 women) agreeing to participate. One hundred thirty-four women did not respond (27.0%) and 67 women declined the study (13.5%). The mothers who participated were similar to non-participants in terms of their age at delivery, husband’s current age, child’s current age, birthweight and gestational age at delivery (Table I).

ITBS/ITED scores at third grade or higher were available on 423 (91.4%) IVF children (308 IVF cycles, 272 women) enrolled in the study. Children who did not have test scores available were most likely either absent on testing day or had inaccurate information in one database which precluded us from matching their records.
The child, parent and IVF characteristics for all subjects with test scores are shown in Table II. The majority of parents were white (97.1% of mothers, 96.0% of fathers), had at least a college education (82.4% mothers, 66.9% fathers) and were still married (88.9%). The mean age of mothers at the time of delivery was 33.9 years (range 25–46 years old). The mean age of the children enrolled in the study was 11.6 years (range 8–17 years). The distribution by gestation type for the 308 cycles (423 children) was 201 singletons (65.3% cycles), 92 twins (29.9% cycles; 178 children) and 15 triplets (4.9% cycles; 44 children). Mean birthweight was 2923.5 g (624–4920 g), and mean gestational age at delivery was 37.1 weeks (range 26–42.4 weeks).

The majority of deliveries resulted from fresh embryo transfers (80.5%) fertilized by standard insemination (73.7%), transferred into the Fallopian tube on Day 1 (ZIFT, 39.9%) or into the uterus on Day 2 or 3 (IVF, 57.8%) and were from autologous oocytes (90.6%).

Figure 1 shows the CT NPR mean (with 95% confidence interval) by grade. For all grades, the mean NPR CT was significantly higher in IVF children ($P < 0.0001$) than the national norm (50th percentile). The same results were found for each subtest NPR with mean NPRs ranging from 70 to 82% (data not shown).

Control ITBS/ITED scores were provided when there were at least 20 children in the matched peer group, so the analysis comparing IVF

<table>
<thead>
<tr>
<th>Table I</th>
<th>Characteristics of participants versus non-participants in the study of achievement test performance in children conceived by IVF.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristic</td>
<td>Participants, $n = 463$ children/335 cycles/295 couples, mean (SD)</td>
</tr>
<tr>
<td>Mother’s age at delivery (years)</td>
<td>33.9 (4.3)</td>
</tr>
<tr>
<td>Father’s current age (years)</td>
<td>48.0 (6.2)</td>
</tr>
<tr>
<td>Child’s current age (years)</td>
<td>11.5 (2.4)</td>
</tr>
<tr>
<td>Infant birthweight (g)</td>
<td>2897.9 (879.0)</td>
</tr>
<tr>
<td>Gestational age (weeks)</td>
<td>37.8 (3.0)</td>
</tr>
</tbody>
</table>

‘Current age’ refers to the age at the time the study was initiated in 2008.

<table>
<thead>
<tr>
<th>Table II</th>
<th>Characteristics of study participants who had ITBS and ITED scores.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents* ($n = 272$)</td>
<td>Children** ($n = 423$)</td>
</tr>
<tr>
<td>Parent’s race</td>
<td>Child’s current age</td>
</tr>
<tr>
<td>Mother—white</td>
<td>264 (97.1%)</td>
</tr>
<tr>
<td>Father—white</td>
<td>261 (96.0%)</td>
</tr>
<tr>
<td>Mother’s education</td>
<td></td>
</tr>
<tr>
<td>Graduate degree</td>
<td>64 (23.5%)</td>
</tr>
<tr>
<td>College/Assoc degree</td>
<td>160 (58.8%)</td>
</tr>
<tr>
<td>GED/high school</td>
<td>48 (17.6%)</td>
</tr>
<tr>
<td>Father’s education</td>
<td>Birthweight (g)</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>67 (24.6%)</td>
</tr>
<tr>
<td>College/Assoc degree</td>
<td>115 (42.3%)</td>
</tr>
<tr>
<td>GED/high school</td>
<td>90 (33.1%)</td>
</tr>
<tr>
<td>Parents divorced</td>
<td>EGA delivery</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Current BMI*</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>Median (IQR)</td>
</tr>
<tr>
<td></td>
<td>Range</td>
</tr>
</tbody>
</table>

GIFT, gamete intra-Fallopian transfer; ZIFT, zygote intra-Fallopian transfer; EGA, estimated gestational age.

*Self-reported data collected from questionnaire.

**Data collected from University of Iowa Hospitals and Clinics IVF database.
children to their matched peers included 372 children. The median number of children that comprised a peer group was \( n = 114 \) (interquartile range 47–369; range 21–1299). IVF children scored higher than their matched peers on all subtests and across all grades. There was a significant grade effect (\( P = 0.011 \) for CT), suggesting that this difference varied according to grade. As shown in Fig. 2, IVF children scored increasingly higher than their peers from third grade to ninth grade (\( P = 0.016 \) for CT), but this trend reversed from ninth grade on.

From the linear mixed model analysis that examined the effect of various factors on ITBS/ITED score (Table III), a significant association was found between mother’s and father’s level of education and RT and RV scores. Mother’s age at delivery also significantly affected scores, with older age associated with higher scores, and this was significant across all five subtests. When we removed children from donor oocyte cycles, this effect was still significant but only for RT, MT and CT. Child’s BMI affected LT and CT significantly, with a higher BMI associated with lower scores. The BMI used was the child’s current BMI, however, and both current and past test scores were included for this analysis. Divorce was negatively associated with CT score. There was no significant association between the type of IVF cycle (fresh versus frozen), type of insemination or length of embryo culture and test scores. Neither birthweight nor gestational age had an effect on test scores.

When looking at ITBS/ITED score by gestation type, singletons scored higher than twins and twins higher than triplets for all five subtests (Fig. 3). These differences were initially statistically significant, but fell below significance after we adjusted for a cluster effect or the effect of scores from children in the same household. The largest z-score difference between singletons and triplets corresponded to a 10.5 percentile difference in RV, and a 7.2 percentile difference in MT. As seen in Fig. 3, the triplets had a large SEM, likely a result of the small sample size in this group.

Data collected from the questionnaire regarding school performance were descriptive only as no control data for these outcomes were available. Out of the 463 IVF children in the study, 5 (1.1%)...
Discussion

This study represents the largest report to date of academic performance in children conceived by IVF. The findings of our study indicate that IVF children perform above average on standardized testing compared with their matched peers. This finding was present across grades 3 through 11 and across all five ITBS/ITED subtests, including RV, RT, LT, MT, and CT. A trend toward lower test scores in multiple gestations was present but was not significant, although small sample size limited this subanalysis. Cryopreservation of embryos, insemination by ICSI and longer embryo culture did not have an effect on test scores. Other factors, including parental level of education, maternal age at the birth, divorce and child’s BMI, did affect scores.

The findings of this study correlate with those from similar previously published studies, which have indicated that the cognitive development of IVF children is normal. However, these previous studies have been limited by small sample size, lack of a control group and lack of a widespread and objective measure of cognition, particularly in children beyond 5 years of age (Cederblad et al., 1996; Olivennes et al., 1996, 1997; Levy-Shiff et al., 1998; Place and Englert, 2003; Bonduelle et al., 2005; Wagenaar et al., 2008a,b; Wennerholm et al., 2009). The strengths of our study include the large sample size, suitable control group, extended length of follow-up to 17 years of age and use of an objective measure of cognition.

This study used test scores from ITBS/ITED to measure cognitive ability in IVF children. In the state of Iowa, USA, virtually all children in grades 3 through 12 are required to undergo annual standardized testing using the Iowa Tests. Batteries of tests in multiple subject areas, the Iowa Tests, require 5–6 h of testing and are given over multiple days. These tests are recognized nationally as valid and reliable measures of academic achievement. We were able to obtain ITBS/ITED scores on 423 (91.4%) of the IVF children we enrolled in the study.

One explanation for our finding of higher test scores in IVF children than their peers may be related to the potentially higher socioeconomic background of IVF children (Zwick and Green, 2007). We attempted to control for this possible confounder by matching each IVF child to a control group of peers by school district. In Iowa, children from the same school district tend to be from similar socioeconomic backgrounds. Although we believe our control group was similar to our IVF group socioeconomically, we do not have any information on our control group to confirm this assumption. We also did not have information from our control group on parental level of education, maternal age at the birth, divorce and child’s BMI, did affect scores.

Another explanation may be a higher level of involvement in the upbringing and education by parents of IVF children, a factor that is difficult to control for. It is also plausible that the higher scores in IVF children could be the result of other unknown confounding factors.

Our finding of lower test scores in children from multiple gestations was interesting and has been reported previously (Tsou et al., 2008). Although the differences were not significant, the trend was consistent across all five subtests. The high rate of prematurity and related sequelae in multiple gestations is unlikely to be the explanation for these findings as gestational age and birthweight were controlled for in our analysis. However, we cannot rule out some other in utero effect of multiple gestation on cognitive development. Another
The lack of long-term follow-up on children conceived through this method is limited, there does not appear to be a negative effect of IVF children on test scores was also interesting and has been reported in other studies as well (Elias et al., 2003; Datar et al., 2004; Li et al., 2008). Mechanisms that have been proposed to explain this relationship have included impaired insulin signaling, low leptin in the brain and obstructive sleep apnea (Amin et al., 2002; Craft and Watson, 2004; Farr et al., 2006). Psychosocial factors within the family may also be responsible for both obesity and lower cognitive function (Lissau and Sorensen, 1994).

The prevalence of ADHD and autism spectrum disorders (ASD) in the IVF children in our study was 7.0 and 0.9%, respectively. These data were parent-reported by questionnaire and descriptive only, as we did not have similar information on children in our control group. However, these rates are similar to the prevalence of parent-reported ADHD and ASD (7.8 and 1.1%, respectively) in school-age children in the USA (Mental Health in the United States, 2005; Kogan et al., 2009).

A limitation of our study was that 40.6% of IVF patients invited to participate either declined or did not respond. Using information from our IVF database on these women, we determined that study participants were similar to non-participants in terms of gestational age at delivery, birthweight and maternal age at delivery, suggesting to us that any sampling bias was likely minimal. However, such bias cannot be ruled out entirely as parents of children with higher scores may have been more likely to participate.

Although we did not have any information on how the children in our control group were conceived, we assumed that they were not the result of IVF in our analysis. Given that 1% of children in our population are conceived through IVF, we would expect this to have diluted the difference between our IVF children and their matched peers, biasing the effect toward the null. Therefore, the effect of IVF on ITBS/ITED performance may have been greater than what we are reporting.

The findings of our study suggest that although long-term follow-up of IVF children is limited, there does not appear to be a negative effect of IVF on cognitive development. The results of this study will be reassuring to patients who are considering IVF but are concerned about the lack of long-term follow-up on children conceived through this technology.

**Acknowledgements**

We would like to thank all of our IVF patients who took the time to participate in our study.

**Funding**

Supported in part by funds from the Mary Ellen Molinaro-Blonigan Research Fund, University of Iowa Carver College of Medicine.

**References**


Place I, Englert Y. A prospective longitudinal study of the physical, psychomotor, and intellectual development of singleton children up to 5 years who were conceived by intracytoplasmic sperm injection compared with children conceived spontaneously and by in-vitro fertilization. *Fertil Steril* 2003; 80:1388–1397.


