CHAPTER 7
The effect of preimplantation genetic screening on neurological, cognitive and behavioural development in 4-year-old children: follow-up of a RCT

Pamela Schendelaar
Karin J. Middelburg
Arend F. Bos
Maas Jan Heineman
Joke H. Kok
Sacha la Bastide-van Gemert
Jorien Seggers
Edwin R. van den Heuvel
Mijna Hadders-Algra

*Human Reproduction* 2013; 28: 1508 – 1518
ABSTRACT

Study question Does embryo biopsy inherent to preimplantation genetic screening (PGS) affect neurological, cognitive and behavioural development of 4-year-old children?

Summary answer PGS does not seem to affect neurological, cognitive and behavioural development of 4-year-old singletons, however, our data suggest that it may be associated with altered neurodevelopment in twins.

What is known already Evidence concerning the safety of PGS on neurodevelopmental outcome in offspring is scarce. The present study provides information on neurodevelopmental, cognitive and behavioural outcome of 4-year-old PGS offspring.

Study design, size, duration A prospective, assessor-blinded follow-up study of children born to women who participated in a multicentre RCT on the effect of IVF with or without PGS.

Participants/materials, setting, methods At 4 years, 49 children (31 singletons, 9 sets of twins) born following IVF with PGS and 64 children (42 singletons, 11 sets of twins) born following IVF without PGS (controls) were assessed (postnatal attrition 18%). Neurological development was evaluated with the standardized, age-specific and sensitive neurological examination according to Hempel, resulting in a neurological optimality score (NOS), a fluency score and the rate of adverse neurological outcome. Primary outcome was the fluency score, as fluency of movements is easily reduced by subtle dysfunction of the brain. Cognitive development was evaluated with the Kaufman Assessment Battery for Children; behavioural development was evaluated with the Child Behavior Checklist. The effect of PGS was analysed with a mixed-effects model.

Main results and the role of chance Based on the intention to treat analysis, neurodevelopmental outcome of PGS children was similar to that of controls. However, additional analyses indicated that PGS affected neurodevelopmental outcome of twins in a different way than that of singletons. The fluency score of singletons born following PGS was similar to that of control singletons (mean values, 95% confidence intervals (CIs): 12.2 [11.5 to 12.8] and 12.2 [11.6 to 12.8] respectively, P = 0.977) that was also true for the other neurodevelopmental parameters. The fluency score of PGS twins was significantly lower than that of control twins (mean values, 95% CIs: 10.6 [9.8 to 11.3] and 12.3 [11.5 to 13.1] respectively, P = 0.001); the same was true for the NOS. In addition, PGS in twins was associated with a higher sequential intelligence quotient score. On the other hand, other neurodevelopmental parameters were similar for PGS twins and control twins. post hoc sample size calculation for the primary outcome parameter, the fluency score, indicated that the study groups, including the subgroups of singletons and twins, were adequately powered.

Limitations, reasons for caution We assessed singletons and twins who contributed to the generalizability of the study. A limitation of our study is the relative small size of our study groups and the selective drop-out in both groups (drop-outs PGS group: higher gestational age; control group: less well-educated parents). These preclude the conclusion that PGS per
se is not associated with neurodevelopmental, cognitive and behavioural problems in singletons and the conclusion that PGS is associated with altered neurodevelopmental outcome in twins.

**Wider implications of the findings** The need for careful long-term monitoring of children born following embryo biopsy remains, as it is still applied in the form of PGD and it is still unknown whether embryo biopsy affects long-term neurodevelopmental outcome.
INTRODUCTION

Preimplantation genetic screening (PGS) was introduced to enhance efficiency in assisted conception. In PGS, a blastomere is aspirated from a cleavage stage embryo via an opening in the zona pellucida that is created by enzymatic digestion or with laser. Next, the copy numbers of several sets of chromosomes are identified by means of fluorescent in situ hybridization and only euploid embryos are transferred to the uterus. In theory, PGS results in improved implantation and increased ongoing pregnancy and live birth rates, as it is assumed that about half of all embryos obtained through IVF are aneuploid. However, a meta-analysis, performed over nine well-designed randomized controlled trials on PGS outcome, reported no evidence for a beneficial effect of PGS on live birth rate. Instead, PGS significantly reduced the live birth rate for women of advanced maternal age.

Despite the invasiveness of the embryo biopsy inherent to PGS, few studies addressed development and growth of children born following PGS. Two research groups reported on developmental outcome of mixed groups of children born following embryo biopsy, i.e. born following PGS or PGD. One series of studies reported on mental, motor, socioemotional and language development in 2-year-old PGD/PGS-children. The studies demonstrated similar outcome in singletons born following PGD/PGS and singletons born following ICSI or naturally conceived singletons. Similar results were found for twins. Another research group reported on health and developmental outcome measured with the Griffiths scale in children born following PGD/PGS up to the age of 4 years. Health and developmental outcome of the PGD/PGS children was similar to that of controls, with the exception of locomotor development, where PGD/PGS children scored significantly lower than controls. The studies did, however, not differentiate between PGD and PGS, although the indication to perform one or another is rather different. In PGD, blastomeres are analysed in couples with a high risk of a genetic disorder. In general, couples who are considered for PGD treatment do not have fertility problems, whereas couples who are considered for PGS treatment do. Underlying subfertility problems may contribute to the occurrence of neurodevelopmental problems in PGS offspring, as subfertility is known to be associated with more obstetrical events and more perinatal adversities.

Previously, our group reported that neurodevelopmental outcome up until 2 years of age in children born following IVF with PGS was largely similar to that of children born following IVF without PGS. However, application of the detailed neurological optimality score (NOS) revealed that children born after IVF with PGS had a somewhat less optimal neurological condition at 2 years than children born after IVF without PGS. This may suggest a less favourable neurological development in children born following PGS.

The literature summarized above indicates that our knowledge on the long-term consequences of embryo biopsy on child development is unclear. We, therefore, decided to reassess the children born following IVF with and without PGS who participated in the follow-up study at 2 years. Reassessment consisted of the evaluation of neurodevelopmental, cognitive and behavioural outcome at the age of 4 years. The study is
part of a prospective, assessor-blinded, follow-up study of children born to women who participated in multicentre RCT on IVF with or without PGS.22

Primary outcome was the neurological condition of the child expressed in terms of fluency of motor behaviour. This is a sensitive measure to detect minor changes in neuromotor development, as fluency of movements is easily affected by subtle dysfunction of the brain. We therefore expected that potential differences between the two groups would be most clearly expressed in the fluency score. Secondary outcome parameters were the NOS, adverse neurological outcome – defined as the occurrence of complex minor neurological dysfunction (MND) or a neurological syndrome – and cognitive function and behaviour. We hypothesized that neurodevelopmental outcome, in particular fluency of movements, of children born following IVF with PGS is less optimal than that of children born after IVF without PGS, as PGS involves embryo biopsy. Specific attention was paid to the effect of PGS on singletons and twins, as Liebaers et al. indicated that embryo biopsy in combination with multiple pregnancy was associated with increased perinatal morbidity and mortality, whereas a similar increased morbidity and mortality was absent after embryo biopsy in singleton pregnancies.23

MATERIAL AND METHODS

Participants

For the present study all children born to women participating in a double blind, RCT on the efficiency of PGS to improve ongoing pregnancy rates after IVF were eligible (ISRCTN76355836).22 Inclusion criteria for participating women were: 1) age 35 to 41 years, 2) no previously failed IVF-cycles and 3) no objections against a possible double embryo transfer (DET). Randomization of women was performed centrally, using a computer program, with minimization for age (35-37 or 38-41 years) and reproductive technology (IVF or ICSI), and with stratification according to study centre [Academic Medical Center (AMC), Amsterdam and University Medical Center Groningen (UMCG), Groningen] prior to the start of the IVF procedures. Group status was revealed to participating parents after 12 weeks of gestation. Details on study design and information concerning IVF treatment procedures and the follow-up program have been reported previously.22,142

Prior to inclusion and randomization, couples were informed on the neurodevelopmental follow-up evaluation as part of the PGS trial. The Dutch Central Committee of Research Involving Human Subjects and the Medical Ethics Committees of the local hospitals approved the protocol of the follow-up study. The children’s parents provided written informed consent for participation of their child(ren) in the follow-up study, including the present follow-up.
**Settings**
Demographic information, for example parity, gestational age, birthweight, neonatal intensive-care unit admission, parental age and educational level, was collected on standardized charts at the first follow-up assessment at 2 weeks.\(^{141}\)

The assessment at 4 years was carried out by trained assessors supervised by a neurodevelopmental expert (M.H-A). The assessors were blind to prenatal and perinatal history, including the mode of conception. All assessments were carried out between October 2009 and November 2010. The children who were conceived at the UMCG were assessed at the Institute of Developmental Neurology at the UMCG, and the children who were conceived at the AMC were assessed at home.

**Measures**

*Neurological development*

Around the time of their fourth birthday, children were assessed with the standardized neurological assessment according to Hempel.\(^{182}\) The Hempel assessment was developed to detect MND at preschool age. Neurological signs are organized in five domains of function: fine motor function, gross motor function, posture and muscle tone, reflexes and visuomotor function (i.e. function of the visual system and eye movements). According to specified criteria each domain is classified as typical or deviant.\(^{155}\)

Children were classified as neurologically normal, simple MND, complex MND or neurologically abnormal. A child is classified as having simple MND when one functional domain (except for the domain reflexes) is scored as deviant. Simple MND is regarded as a non-optimal, yet normal form of brain function.\(^{90}\) A child is classified as having complex MND when more than one domain is scored as deviant. Complex MND represents the clinically relevant form of MND, as it is associated with prenatal and perinatal adversities and with learning and behavioural disorders.\(^{90,157}\) Neurologically abnormal implies the presence of a distinct neurological syndrome such as cerebral palsy. Neurologically normal implies the absence of neurological dysfunction and is scored when no domains are deviant or only the domain of reflexes is deviant. We considered complex MND and neurologically abnormal as adverse neurological outcome.

Additionally, the outcome of the Hempel assessment was expressed in a NOS. The NOS consists of 56 items (range 0-56), for which an optimal condition is defined. The total score results from the sum of the items that fulfil the criteria for optimality. Higher scores represent better performance. It is important to realize that there is a conceptual difference between normality and optimality because the range for optimal behaviour is narrower than that for normal behaviour.\(^{160}\) This implies that the NOS is sensitive to minor changes, including those lying within the normal range. The fluency score (range 0-15) is a subscore of the NOS that evaluates the fluency of motor behaviour. The fluency score is the most sensitive measure to detect minimal changes in neuromotor development because subtle dysfunction of the nervous system is most easily expressed in a reduction in the fluency of
movements. The inter-rater reliability of the Hempel assessment is satisfactory [reliability per item: $\kappa = 0.62-1.00$ (mean 0.93)] and its construct validity is good.\(^{155,163-165}\)

**Cognitive development**

Cognitive development was evaluated by means of the Kaufman Assessment Battery for Children, second edition (K-ABC-II).\(^{166}\) The K-ABC-II is an individually administered standardized clinical instrument that measures cognitive and processing abilities in children and adolescents (3 – 18 years). Cognitive and processing abilities are expressed in a total intelligence quotient (IQ) score (the so-called Fluid-Crystallized Index) and four IQ scale scores: 1) a sequential processing IQ that reflects the short-term memory of a child, 2) a simultaneous processing IQ, representing the spatial aptitude of a child, 3) a learning ability IQ, representing the long-term memory capacity of a child and 4) a knowledge IQ. All raw scores are normalized into global scores (mean: 100; standard deviation [SD]: 15). The K-ABC-II has been used for children with different social backgrounds or ethnic differences without critical effects on test-scores. Reliability and validity of the K-ABC-II are good.\(^{166}\) American norms were applied because Dutch norms are lacking.

**Behavioural development**

Behavioural development was evaluated by means of the Dutch version of the Child Behavior Checklist (CBCL).\(^{167,168}\) The CBCL is a parental questionnaire designed to identify emotional and behavioural problems in children aged 1½ to 5 years. Questions on the CBCL are grouped into the following problem scales; emotionally reactive, anxious/depressed, somatic complaints, withdrawn, sleep problems, attention problems and aggressive behaviour. The first four of these scales together form the internalizing scale and the latter two form the externalizing scale. The sum of all questions determines the total problem scale. Raw scores are normalized into T-scores (mean: 50; SD: 10). Higher scores represent more problematic behaviour. Reliability and validity of the CBCL are good.\(^{167}\)

**Statistical analysis**

Fisher’s exact tests and Mann-Whitney U-tests were applied to compare demographic characteristics between the PGS group and the control group at parent level and at child level separately for singletons and twins.

A linear mixed-effects model, with a random effect for mother to model the possible correlation between twins, was performed to analyse the potential effect of PGS for twins and singletons separately on the primary outcome measure fluency score and the secondary outcome measures (NOS, the various IQ scores and the three CBCL scores). A type III t-test with Satterthwaite’s degrees of freedom\(^{173}\) was used for the specific contrasts. The effect of PGS on the additional secondary outcomes MND and adverse neurological outcome was estimated for singletons and twins separately using generalized estimating equations with a logit link function and a binomial distribution. The Wald test was applied
for the specific contrasts. For the numerical outcome measures, all analyses were adjusted for centre and assessor, whereas for the binary outcome measures (MND, adverse neurological outcome), all analyses were adjusted for centre only.

**FIGURE I.** Flow chart on eligibility and participation of children and parents in the PGS RCT and follow-up study.
A post hoc power analysis was conducted using a non-central t-test to estimate the minimal detectable effect size for singletons and twins separately or together, depending on the outcome of the test. The type I error rate and power were selected at 5% and 80% respectively.

All analyses were performed according to the ‘intention to treat’ (ITT) principle in which the four children born to couples who conceived naturally and the one child that was conceived via intra-uterine insemination (IUI) were taken into account. To explore potential effects of natural conception and IUI, we performed sensitivity analyses in which we repeated all analyses with the exclusion of the four naturally conceived children and the one IUI child. The sensitivity analysis also included an exploration of a potential interaction effect of ICSI with IVF for singletons and twins. Finally, we adjusted for gestational age in the primary analyses to investigate whether or not the effect of PGS was mediated by gestational age, as it is known that gestational age is an important predictor for neurological outcome.

All analyses were performed with the Statistical Analysis System software version 9.2 or the Predictive Analytics SoftWare Statistics, version 18. In all analyses probability values of 5% or less were considered significant.

RESULTS

Participation and demographic characteristics
Between May 2003 and November 2005, 408 women were included in the randomized trial. The trial resulted in 52 ongoing PGS pregnancies and 74 control pregnancies of which respectively 49 (39 singletons, 10 sets of twins) and 71 (57 singletons, 14 sets of twins) pregnancies resulted in live births (Figure I). A set of twins died immediately postpartum due to immature birth. Four couples could not be invited for follow-up due to the withdrawal of informed consent during treatments. The reason for withdrawal was in most cases the stress caused by the blinding of couples for treatment allocation. Eventually, 47 PGS couples and 68 control couples were eligible for follow-up evaluation. At 4 years of age, 40 (85%) PGS couples with 49 children (31 singletons, 9 sets of twins) and 53 (78%) control couples with 64 children (42 singletons, 11 sets of twins) participated in the study (attrition rate of 18% with regard to baseline, Figure I). Overall, background factors of non-participants were similar to those of participants, with the exception of selective drop-out of less well-educated control parents (mothers: \( P = 0.006 \); fathers: \( P = 0.040 \)), and PGS children with a higher gestational age (\( P = 0.040 \)).

Demographic characteristics of parents and children of both study groups are shown in Table I. At parent level, demographic characteristics of the groups were similar, except for Caesarean section: PGS children were less often delivered by Caesarean section (\( P = 0.039 \)). At child level, all demographic characteristics of the groups were similar, also in the subgroups of singletons and twins, except for gestational age: children of the PGS group...
### TABLE I. Characteristics of parents and infants in the PGS follow-up study.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Couples with ≥ 1 live birth after PGS</th>
<th>Children born after PGS</th>
<th>Control couples</th>
<th>Control children</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 40</td>
<td>n = 49</td>
<td>n = 31</td>
<td>n = 18</td>
</tr>
<tr>
<td>Parental characteristics</td>
<td></td>
<td></td>
<td>n = 53</td>
<td>n = 64</td>
</tr>
<tr>
<td>Maternal age at conception in years, median (range)</td>
<td>37.3 (35.2 - 40.9)</td>
<td>37.9 (35.3 - 41.3)</td>
<td>31 (39)</td>
<td>27 (58)</td>
</tr>
<tr>
<td>Education level mother (higha), n (%)</td>
<td>21 (55)</td>
<td>22 (55)</td>
<td>31 (99)</td>
<td>27 (51)</td>
</tr>
<tr>
<td>Education level father (higha), n (%)</td>
<td>21 (53)</td>
<td>22 (55)</td>
<td>31 (99)</td>
<td>27 (51)</td>
</tr>
<tr>
<td>Time to pregnancy (years), median (range)</td>
<td>3.9 (1.1 - 12.4)</td>
<td>4.1 (0.7 - 15.9)</td>
<td>3 (9)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Smoking during pregnancy, n (%)</td>
<td>3 (9)</td>
<td>4 (11)</td>
<td>1 (3)</td>
<td>0</td>
</tr>
<tr>
<td>Alcohol consumption during pregnancy, n (%)</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>0</td>
</tr>
<tr>
<td>Conception method</td>
<td></td>
<td></td>
<td>n = 53</td>
<td>n = 64</td>
</tr>
<tr>
<td>IVF performed, n (%)</td>
<td>24 (60)</td>
<td>29 (60)</td>
<td>22 (42)</td>
<td>22 (42)</td>
</tr>
<tr>
<td>ICSI performed, n (%)</td>
<td>13 (33)</td>
<td>22 (42)</td>
<td>22 (42)</td>
<td>22 (42)</td>
</tr>
<tr>
<td>IUI performed, n (%)</td>
<td>1 (3)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Natural, n (%)</td>
<td>2 (5)</td>
<td>2 (4)</td>
<td>2 (4)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>PGS treatment centre (UMCG), n (%)</td>
<td>22 (53)</td>
<td>28 (53)</td>
<td>28 (53)</td>
<td>28 (53)</td>
</tr>
<tr>
<td>Child characteristics</td>
<td></td>
<td></td>
<td>n = 53</td>
<td>n = 64</td>
</tr>
<tr>
<td>Twins, n (%)</td>
<td>18 (37)</td>
<td>19 (34)</td>
<td>18 (32)</td>
<td>14 (22)</td>
</tr>
<tr>
<td>Male gender, n (%)</td>
<td>26 (53)</td>
<td>31 (59)</td>
<td>26 (53)</td>
<td>26 (53)</td>
</tr>
<tr>
<td>Corrected age at examination at 4 years of age (months), median (range)</td>
<td>50.9 (47.8 - 67.9)</td>
<td>50.9 (47.8 - 67.9)</td>
<td>51.1 (48.4 - 59.8)</td>
<td></td>
</tr>
<tr>
<td>Birth characteristics</td>
<td></td>
<td></td>
<td>n = 53</td>
<td>n = 64</td>
</tr>
<tr>
<td>First born, n (%)</td>
<td>24 (60)</td>
<td>29 (60)</td>
<td>22 (42)</td>
<td>22 (42)</td>
</tr>
<tr>
<td>Gestational age (weeks), median (range)</td>
<td>39.4 (32.0 - 42.0)</td>
<td>37.7 (36.9 - 46.0)</td>
<td>37.1 (32.0 - 39.3)</td>
<td>39.1 (30.7 - 41.6)</td>
</tr>
<tr>
<td>Preterm birth (&lt; 37 weeks), n (%)</td>
<td>5 (13)</td>
<td>9 (17)</td>
<td>2 (5)</td>
<td>14 (30)</td>
</tr>
<tr>
<td>Birth weight (grams), mean (SD)</td>
<td>3219 (712)</td>
<td>3861 (543)</td>
<td>3613 (543)</td>
<td>3073 (657)</td>
</tr>
<tr>
<td>Low birth weight (&lt; 2500 grams), n (%)</td>
<td>9 (23)</td>
<td>7 (12)</td>
<td>9 (17)</td>
<td>11 (22)</td>
</tr>
<tr>
<td>Small for gestational age, n (%)</td>
<td>2 (5)</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Caesarean section, n (%)</td>
<td>7 (18)</td>
<td>14 (18)</td>
<td>9 (17)</td>
<td>14 (27)</td>
</tr>
</tbody>
</table>

Note: Fisher's exact tests, Student's t-test and Mann-Whitney U-tests were applied to compare demographic characteristics between the PGS group and the control group at parent level and at child level separately for singletons and twins. * P < 0.05, ** P < 0.1.

PGS = preimplantation genetic screening; IVF = in vitro fertilization; ICSI = intracytoplasmic sperm injection; IUI = intrauterine insemination; UMCG = University Medical Center Groningen.

a University education or vocational colleges.
b Treatment was converted to IUI in case of poor follicle growth.
c Data were analysed according to intention to treat.
d Birthweight for gestational age is < -2 standard deviations compared with the Dutch reference population (Dutch reference tables, perinatal registration Netherlands).
e Missing data in two groups: time to pregnancy n=1, smoking during pregnancy n=9, alcohol consumption during pregnancy n=12, Caesarean section n=1, Apgar score 5 min < 7 n=8.

f Birthweight for gestational age is < -2 standard deviations compared with the Dutch reference population (Dutch reference tables, perinatal registration Netherlands).
had a higher gestational age at birth ($P = 0.027$) than children of the control group. In the subgroups, PGS singletons had a higher gestational age at birth ($P = 0.005$) than controls; gestational age of twins was similar.

The mixed-effects model indicated that PGS affected neurodevelopmental outcome of twins in a different way than that of singletons. Therefore, the results on outcome are presented for singletons and twins separately.

**Neurological development at 4 years**

Note that the results on neurological outcome in clinical terms of adverse neurological outcome are presented first. Thereafter, the results of the primary outcome parameter, the fluency score, and those of the NOS are presented.

Based on the ITT analysis, neurological outcome of PGS children was similar to that of controls. However, the mixed-effects models indicated that neurological outcome differed for singletons and twins (Table II). The rate of adverse neurological outcome in PGS children and controls was similar in singletons (PGS: n=3 (10%); controls: n=10 (24%), $P = 0.145$) and twins (PGS: n=6 (33%); controls: n=3 (14%), $P = 0.116$). However, one member of a PGS twin was diagnosed with cerebral palsy, whereas no members of control twins had cerebral palsy. The fluency score of singletons born following PGS was similar to that of control singletons ($P = 0.977$). However, PGS twins had significantly lower fluency scores than control twins [mean values, 95% confidence intervals (CIs): 10.6 (9.8 to 11.3) and 12.3 (11.5 to 13.1) respectively, $P = 0.001$]. Similarly, the NOS did not differ between PGS singletons and controls ($P = 0.548$) but PGS twins had a significantly lower NOS than control twins [mean values, 95% CIs: 44.8 (42.8 to 46.9) and 48.8 (46.8 to 50.9) respectively, $P = 0.005$].

After adjusting for gestational age in the analyses concerning the fluency score and the NOS, the differences between the groups remained statistically significant ($P = 0.001$ and $P = 0.005$, respectively). Exclusion of the PGS child with cerebral palsy did not alter the conclusions about the group differences in fluency score and NOS. Also the exclusion of the four naturally conceived children and the one IUI child did not alter the conclusions (data not presented). Moreover, the effect of PGS on the neurodevelopmental outcome measures did not differ for ICSI and IVF singletons and twins (data not presented).

**Cognitive development at 4 years**

Total IQ scores of all children were in the normal range, except for one singleton of the control group who had a total IQ score of 82. Based on the ITT analysis, cognitive development of PGS children was similar to that of controls but again differences were found for the effect of PGS in singletons and twins (Table II). The total IQ scores of PGS singletons did not differ from those of control singletons ($P = 0.666$). Moreover, no differences were found in the four IQ scale scores between PGS singletons and controls, except for the learning IQ score that turned out to be significantly lower among PGS singletons when compared to control singletons ($P = 0.049$) (Table II). However, after adjusting for gestational age in the analyses, the difference disappeared ($P = 0.066$). The
total IQ scores of PGS twins did not differ from those of control twins ($P = 0.104$). Moreover, no differences were found in the four IQ scale scores between PGS twins and control twins, except for the sequential IQ score that turned out to be significantly higher among PGS twins compared to control twins ($P = 0.027$), also after adjusting for gestational age ($P = 0.028$). Exclusion of the PGS child with cerebral palsy did not alter the conclusions on group differences in cognitive outcomes. Also when the one IUI child and the four naturally conceived children were excluded, the conclusions did not change (data not presented). Moreover, the effect of PGS on the cognitive outcome measures did not differ for ICSI and IVF singletons and twins (data not presented).

**Behavioural development at 4 years**

Based on the ITT analysis, behavioural development of PGS children was similar to that of controls. No differences between the groups were found in the total problem score in singletons ($P = 0.357$) and twins ($P = 0.983$) or in internalizing and externalizing behaviour (Table II).

Again, the sensitivity analyses did not demonstrate alternative results when the one IUI child and the four naturally conceived children were excluded (data not presented). Moreover, we did not detect heterogeneity between the effect of PGS for ICSI and IVF singletons and twins on the behavioural outcome measures (data not presented). Also exclusion of the PGS child with cerebral palsy did not alter the conclusions about the group differences in behavioural outcomes.

**Post hoc power analysis**

We performed a post hoc sample size calculation for the primary outcome parameter, the fluency score, for singletons and twins separately. The post hoc power analysis indicated that the current sample size could detect a difference of 1.0 and 1.4 on the fluency score with 80% power and 5% significance level for singletons and twins respectively. This means that our study groups are adequately powered for the primary outcome parameter.

Also, our singleton and twin groups are sufficiently powered for the NOS with a minimal detectable effect size of 2.6 and 3.8 respectively. However, no relevant clinical effect sizes could be detected with the current sample size for the outcome adverse neurological outcome (odds ratios of more than 7.3 and 8.2 for singletons and twins, respectively). For the cognitive outcomes, an effect size of less than seven points with 80% or more power was detectable for singletons, with the exception of the simultaneous IQ score and the knowledge IQ score, where the power was equal to 75% and 63%, respectively. For twins, an effect size of less than 11 points was detectable with more than 80% power for the learning IQ score and the total IQ score; for the simultaneous IQ score, the knowledge IQ score and the total IQ score the power was determined at 75%, 70% and 57%, respectively. For the behavioural outcomes the study groups were adequately powered.
The effect of preimplantation genetic screening on neurological, cognitive and behavioural development in 4-year-old children: follow-up of a RCT

### Neurological development

<table>
<thead>
<tr>
<th></th>
<th>PGS singletons (n = 30)</th>
<th>Control singletons (n = 41)</th>
<th>P - value</th>
<th>PGS twins (n = 16)</th>
<th>Control twins (n = 22)</th>
<th>P - value</th>
<th>P - value interaction effect singleton*twins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluency score, mean [CI]</td>
<td>12.2 [11.5 ; 12.8]</td>
<td>12.2 [11.6 ; 12.8]</td>
<td>0.977</td>
<td>10.6 [9.8 ; 11.3]</td>
<td>12.3 [11.5 ; 13.1]</td>
<td>0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>Neurological optimality score (NOS), mean [CI]</td>
<td>49.3 [47.6 ; 50.9]</td>
<td>48.7 [47.1 ; 50.3]</td>
<td>0.548</td>
<td>44.8 [42.8 ; 46.9]</td>
<td>48.8 [46.8 ; 50.9]</td>
<td>0.005</td>
<td>0.007</td>
</tr>
</tbody>
</table>

**PGS singletons (n = 30) Control singletons (n = 42) P - value PGS twins (n = 16) Control twins (n = 22) P - value P - value interaction effect singleton*twins**

| Minor neurological dysfunction (MND)a, n (%) | 8 (27) | 16 (38) | 0.312 | 6 (33) | 8 (36) | 0.859 | 0.615 |
| Adverse neurological outcomeb, n (%) | 3 (10) | 10 (24) | 0.145 | 6 (33) | 3 (14) | 0.116 | 0.036 |

### Cognitive development

<table>
<thead>
<tr>
<th></th>
<th>PGS singletons (n = 30)</th>
<th>Control singletons (n = 42)</th>
<th>P - value</th>
<th>PGS twins (n = 16)</th>
<th>Control twins (n = 22)</th>
<th>P - value</th>
<th>P - value interaction effect singleton*twins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total IQ, mean [CI]</td>
<td>113.4 [109.7 ; 117.4]</td>
<td>114.4 [110.7 ; 118.1]</td>
<td>0.666</td>
<td>109.8 [104.2 ; 115.5]</td>
<td>104.1 [98.8 ; 109.3]</td>
<td>0.104</td>
<td>0.105</td>
</tr>
<tr>
<td>Sequential IQ, mean [CI]</td>
<td>98.0 [93.7 ; 102.3]</td>
<td>100.7 [96.6 ; 104.9]</td>
<td>0.259</td>
<td>103.5 [96.9 ; 110.2]</td>
<td>94.2 [88.1 ; 100.3]</td>
<td>0.027</td>
<td>0.014</td>
</tr>
<tr>
<td>Learning IQ, mean [CI]</td>
<td>98.7 [94.3 ; 103.1]</td>
<td>103.5 [99.2 ; 107.8]</td>
<td>0.049</td>
<td>96.3 [90.7 ; 101.9]</td>
<td>95.1 [89.7 ; 100.4]</td>
<td>0.710</td>
<td>0.143</td>
</tr>
<tr>
<td>Simultaneous IQ, mean [CI]</td>
<td>123.2 [118.4 ; 128.0]</td>
<td>121.2 [116.6 ; 125.8]</td>
<td>0.447</td>
<td>116.0 [109.6 ; 123.1]</td>
<td>113.2 [106.7 ; 119.7]</td>
<td>0.519</td>
<td>0.876</td>
</tr>
<tr>
<td>Knowledge IQ, mean [CI]</td>
<td>118.6 [113.2 ; 124.0]</td>
<td>116.0 [110.8 ; 121.2]</td>
<td>0.389</td>
<td>111.6 [103.5 ; 119.8]</td>
<td>106.8 [99.3 ; 114.3]</td>
<td>0.341</td>
<td>0.706</td>
</tr>
</tbody>
</table>

### Behavioural development

<table>
<thead>
<tr>
<th></th>
<th>PGS singletons (n = 30)</th>
<th>Control singletons (n = 41)</th>
<th>P - value</th>
<th>PGS twins (n = 16)</th>
<th>Control twins (n = 22)</th>
<th>P - value</th>
<th>P - value interaction effect singleton*twins</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBCL total score, mean [CI]</td>
<td>45.7 [41.8 ; 49.6]</td>
<td>47.7 [44.0 ; 51.5]</td>
<td>0.357</td>
<td>46.1 [40.6 ; 51.5]</td>
<td>46.1 [40.8 ; 51.5]</td>
<td>0.983</td>
<td>0.641</td>
</tr>
<tr>
<td>CBCL internalizing score, mean [CI]</td>
<td>47.9 [41.7 ; 52.1]</td>
<td>49.0 [44.9 ; 53.1]</td>
<td>0.630</td>
<td>47.5 [41.8 ; 53.2]</td>
<td>47.4 [41.7 ; 53.1]</td>
<td>0.981</td>
<td>0.779</td>
</tr>
<tr>
<td>CBCL externalizing score, mean [CI]</td>
<td>47.5 [44.0 ; 51.1]</td>
<td>48.4 [44.9 ; 51.8]</td>
<td>0.684</td>
<td>47.8 [42.8 ; 52.8]</td>
<td>47.3 [42.4 ; 52.3]</td>
<td>0.902</td>
<td>0.750</td>
</tr>
</tbody>
</table>

**Note that the numerical outcomes are adjusted for centre (UMCG, Groningen, Groningen or AMC, Amsterdam) and assessor and that the binary outcome measures are adjusted for centre only.**

IQ = intelligence quotient, CBCL = Child Behaviour Checklist, CI = confidence interval.

*a MND = simple and complex minor neurological dysfunction.

*b Adverse neurological outcome = complex MND or worse (such as cerebral palsy).
Overall, the post hoc analyses indicate that the study sample sizes are large enough to detect relevant differences between PGS and controls singletons and twins separately for the primary outcome parameter and for most secondary outcome parameters.

DISCUSSION

The present study did not demonstrate statistically significant differences in neurological, cognitive and behavioural outcome at 4 years in singletons born following IVF with PGS – involving embryo biopsy – and those born following IVF without PGS. However, we demonstrated statistically significant differences between twins born following IVF with PGS and twins born following IVF without PGS. Our findings on singletons are similar to those of the two other groups that reported on developmental outcome of singletons born after embryo biopsy.\textsuperscript{229,232,233} However, it should be noted that these studies did not differentiate between PGD and PGS, although the indication to perform one or another is rather different. Our finding that PGS affected developmental outcome of twins does not, however, correspond to studies of others.\textsuperscript{228,229} It matches, however, to some extent the study of Liebaers et al., who found higher rates of prematurity and low birthweight in PGD/PGS multiples than in ICSI multiples and more perinatal deaths in post PGD/PGS multiple pregnancies than in post ICSI multiple pregnancies, whereas outcomes were similar for PGD/PGS and ICSI singletons.\textsuperscript{231} Yet, in a larger analysis by Desmyttere et al., including the analyses reported previously by Liebaers et al., no differences in prematurity, low birthweight and perinatal deaths were found between PGD/PGS multiples in comparison with ICSI multiples.\textsuperscript{234,235} This means that the effects of PGD/PGS are inconsistent. It is also good to realize that our results do not necessarily suggest an overall adverse effect of PGS on outcome of twins, as PGS was associated with a negative effect on neuromotor condition and a positive one on sequential processing, an indicator of short-term memory. Our results rather suggest that PGS affects neurodevelopmental outcome of twins in a different way than that of singletons.

We previously reported on outcome of the same group of children exclusively born after PGS. At 2 years of age, we found that children born after IVF with PGS had, on average, an approximately two points lower NOS than controls.\textsuperscript{142} At 4 years of age we found a similar effect in PGS twins, who scored approximately two points lower on the fluency score and approximately four points lower on the NOS when compared to the controls. However, such an effect was not present in singletons at 4 years because the mean difference reduced to less than one point. At the 2 years of follow-up singletons and twins were analysed together, which means that we do not know whether the effect of PGS at that time also could be attributed to the effect on twins. The current findings, however, indicate that PGS may affect neurological condition of twins but not that of singletons. In addition, the appearance of a clearer effect of PGS at 4 years may be related to the characteristics of the developing brain. It is well known that with increasing age, children may grow into a
deficit, i.e. that neurological dysfunctions first emerge when the neural circuitries subserving specific functions become functionally active.90

**Strengths and limitations**

To our knowledge, this is the first follow-up study of a RCT on PGS in which children were followed as long as 4 years. The random assignment of parents resulted in a strong resemblance of the two groups with respect to most background variables so that comparisons between both groups primarily reflect the effect of PGS. Additional strengths with regard to study design are the blinding of the assessors to the mode of conception and the prospective design of the study. Parents were invited to participate before pregnancy so that potential selection bias based on the child’s development or health was reduced. Another strength of the present study is the examination of both singletons and twins, as it is known that being a member of a multiple is associated with an increased risk of developmental problems, irrespective of assisted conception.179 Moreover, the examination of both singletons and twins contributes to the generalizability of the study. It may be argued that another strength is that we included in both PGS and control groups children born after IVF and ICSI. IVF and ICSI are both applied whenever PGS is indicated or not. This contrasts with PGD treatment, in which IVF with ICSI is virtually always used. However, it could be argued that ICSI substantially differs from IVF alone. We therefore applied not only minimization for reproductive technique (IVF or ICSI) during enrolment but also a sensitivity analysis, of which the results indicated that the presence or absence of ICSI did not modify a potential effect of PGS.

A major strength is the application of sensitive and age-specific measurements to assess neurological condition. The strength of the Hempel assessment is illustrated by the study of Bouwstra et al., in which a negative effect of neonatal trans-fatty acid status on neurodevelopmental outcome was demonstrated with the assessment according to Hempel but not with the Bayley’s Scale of Infant Development.164 Although subtle differences in neurological outcome, such as a few points reduction in the NOS or fluency score, may not have clinical relevance for individuals, minor differences in neurodevelopmental outcome may affect a particular subgroup of society, such as subfertile people, a group that is steadily expanding.236

The major limitation of the present study is the relatively small sample size. Power calculation was based on the number of women needed to detect a certain increase in the cumulative ongoing pregnancy rate.22 The unforeseen effect of lower pregnancy rates after PGS further reduced the number of children available for follow-up. We performed a post hoc sample size calculation for the primary outcome parameter, the fluency score, which indicated that our study groups were adequately powered also for the subgroups of singletons and twins. For several other secondary outcome measures the minimal detectable effect sizes were still acceptable, and only for the outcome parameter adverse neurological dysfunction was the study too small. This means that in general our study groups are adequately powered because relevant differences could have been detected.
However, it should be noted that there was a selective loss of less well-educated parents in the control group and a selective loss of children with higher gestational age in the PGS group. Perhaps parents of infants with a lower gestational age are more concerned about their child’s development and therefore more willing to participate in a detailed neurodevelopmental assessment. Presumably, the selective loss of less well-educated parents in the control group does not adequately explain the different neurodevelopmental outcome of PGS twins when compared to control twins, as the selective loss was present in twins and singletons, where the selective loss in the singletons was not associated with developmental differences between PGS children and controls. Moreover, we explored the effects of confounding variables on our outcome measures in an explorative analysis to make sure that the effects found were not mediated by these confounders. This analysis revealed among others no statistically significant effects of parental educational level or gestational age.

The groups, especially the twin groups, were relatively small, implying that the distribution of scores should be taken into account. For the NOS and fluency, Pearson’s residuals did not demonstrate a violation of the assumption of normality, but one twin with cerebral palsy had a score of more than 3 SDs away from the predicted value. This twin can be defined as an outlier in our data. However, excluding this twin from the analyses did not change the conclusions on the significance of the effect of PGS on twins for the NOS and fluency.

The participating children of both groups turned out to have a relatively high IQ. An explanation for this finding may be the relatively large proportion of highly educated parents participating in our study, as it is well known that cognitive abilities of children are positively associated with their parents education level.237 This means that our findings have to be interpreted with caution.

The application of PGS in women of advanced age is debated.23 The European Society of Human Reproduction and Embryology PGD Consortium recommends to apply PGS only in the context of properly constructed trials.24 Even though the use of PGS currently seems to be decreasing,238 embryo biopsy is still applied, for example, in the form of PGD.236 Another issue of concern is the application of DET in IVF with embryo biopsy. DET in IVF with embryo biopsy may entail an accumulation of several risks for offspring, as DET per se increases the risk of twin pregnancy,239-240 and perinatal mortality241 when compared to single embryo transfer in IVF.

In conclusion, the results of the present study suggest that neurological, cognitive and behavioural outcome at 4 years in singletons born following IVF with PGS is similar to that of singletons born after IVF without PGS. In contrast, PGS does seem to affect neurodevelopmental outcome of 4-year-old twins. PGS in twins was associated with a difference in brain function at 4 years, i.e. with a negative effect on neuromotor condition and a positive one on sequential processing. This may point to the possibility that the embryo biopsy inherent to PGS is associated with differences in brain function at a later age.
Hence, the need for careful monitoring of children born following embryo biopsy remains, for it is not yet known whether there are any late consequences of embryo biopsy on neurodevelopmental outcome.

ACKNOWLEDGEMENTS

We thank participated parents and children for their cooperation and enthusiasm during the assessments; Maaike Haadsma, M.D., Ph.D., for her help in including the participants; Rosan Aapkes, B.Sc., Elise Bennik, M.Sc., Hanneke Geut, M.D., Marjolein Jongbloed-Pereboom, M.Sc. and Bertine de Vries, M.Sc., for support in collecting the data; Michiel Schrier, M.Sc., Linze Dijkstra B.Sc., and Loes de Weerd for technical assistance.