Developmental outcome of 9-year-old children born after PGS: Follow-up of a randomized trial

D. Kuiper
A. Bennema
S. la Bastide-van Gemert
J. Seggers
P. Schendelaar
S. Mastenbroek
A. Hoek
M.J. Heineman
T.J. Roseboom
J.H. Kok M.
Hadders-Algra

Human Reproduction 2018; 33: 147-55
Abstract

Study question
Does day-3 cleavage-stage preimplantation genetic screening (PGS) affect neurodevelopment of 9-year-old in vitro fertilization (IVF) offspring?

Summary answer
We did not find evidence of adverse consequences of day-3 cleavage-stage PGS on neurodevelopment of 9-year-old IVF offspring. Children born after IVF with and without PGS often had a non-optimal neurological condition.

What is known already
Knowledge on long-term sequelae for development and health of children born following PGS is lacking. This is striking as evidence accumulates that IVF itself is associated with increased risk for impaired health and development in the offspring.

Study design, size, duration
A prospective, assessor-blinded, multicentre, follow-up study evaluating development and health of 9-year-old IVF children born to women who were randomly assigned to IVF with PGS (PGS group) or without PGS (control group). The follow-up examination at 9 years took place between March 2014 and May 2016.

Participants/materials, setting, methods
In total, 408 women were included and randomly assigned to IVF with or without day-3 cleavage-stage PGS. This resulted in 52 ongoing pregnancies in the PGS group and 74 in the control group. In the PGS group 59 children were born alive; in the control group 85. At the age of 9 years 43 children born after PGS and 56 control children participated in the study. Our primary outcome was the neurological optimality score, a sensitive measure of neurological condition assessed with a standardized, age-specific test (Touwen test). Secondary outcomes were adverse neurological condition (neurologically abnormal and the complex form of minor neurological dysfunction), cognitive development (intelligence quotient and specific domains); behaviour (parental and teacher’s questionnaires), blood pressure and anthropometrics.

Main results and the role of chance
Neurodevelopmental outcome of PGS children did not differ from that of controls; the neurological optimality scores (mean values [95% CI]: PGS children 51.5 [49.3; 53.7], control children 53.1 [50.5; 55.7]) were not significantly different. The prevalences of adverse neurological outcome (in all but one child implying the presence of the complex form of minor neurological dysfunction) did not differ between the groups (PGS group 17/43 [40%], control group 19/56 [34%]), implying also that the prevalence of complex minor neurological dysfunction in both groups was rather high. Also intelligence quotient scores of the two groups were not significantly different (PGS group 114 [108; 120]; Control group 117 [109; 125]). Behaviour, blood pressure and anthropometrics of both groups did not differ. Mean blood pressures of both groups were above the 60th percentile.

Limitations, reasons for caution
The power analysis of the study was not based on the number of children needed for the follow-up study, but on the number of women who were needed to detect an increase in ongoing pregnancy rates after PGS. In addition, our study evaluated embryo biopsy in the form of PGS at cleavage stage (day-3 embryo biopsy), while currently PGS at blastocyst stage (day-5 embryo biopsy) is recommended and increasingly being used.

Wider implications of the findings
Our findings indicate that PGS in cleavage stage embryo’s is not associated with adverse effects on neurological, cognitive and behavioural development, blood pressure and anthropometrics at 9 years. This is a reassuring finding as embryo biopsy in the forms of PGS and preimplantation genetic diagnosis is increasingly applied. However, both groups of IVF offspring showed high prevalences of the clinically relevant form of minor neurological dysfunction, which is a point of concern for the IVF-community. In addition, our study confirms findings of others that IVF offspring may be at risk of unfavourable cardiovascular outcome. These findings are alarming and highlight the importance of research on the underlying mechanisms of unfavourable neurodevelopmental and cardiovascular outcomes of IVF offspring.

Study funding/competing interest(s) (optional): The randomized controlled trial was financially supported by the Organization for Health Research and Development (ZonMw), The Netherlands (grant number 945-03-013). The follow-up was financially supported by the University Medical Center Groningen (grant number: 754510), the Cornelia Foundation, the graduate schools BCN and Share, Groningen, The Netherlands. The sponsors of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report. There are no conflicts of interest.

Trial registration number: ISRCTN76355836.
Introduction

Pregnancy rates after IVF are still lower than desired. To improve the effectiveness of IVF, PGS has been developed. Worldwide, PGS is increasingly used in IVF treatment. Yet, little is known about development and health of children born following PGS.

For many years, PGS consisted of blastomere aspiration from a Day-3 cleavage stage embryo by opening the zona pellucida and blastomere screening with fluorescence in situ hybridization for aneuploidies. Only euploid embryos qualified for embryo transfer. In theory, this should have resulted in improved pregnancy rates, because chromosomal aneuploidy in embryos is considered as one of the main reasons for poor pregnancy rates after IVF. However, multiple randomized controlled trials showed that PGS did not increase, but reduced, pregnancy and live birth rates after IVF in women of advanced maternal age. As a consequence, Day-3 cleavage-stage PGS was no longer recommended.

But times have changed and currently the use of PGS is steadily increasing. Two factors facilitated this change. First, comprehensive chromosome screening allows screening for aneuploidy of all 24 chromosomes, rather than the 5-to-9 chromosomes in the beginning of PGS. Second, a Day-5 blastocyst biopsy seems to be less harmful than Day-3 cleavage-stage biopsy. From the end of the first decade of the 21st century, this technique has been gradually implemented in clinical practice. However, Day-3 cleavage stage embryo biopsy is still carried out. The reasons that the Day-5 biopsy has not been fully implemented is two-fold. First, not all embryos reach blastocyst stage; second, not all fertility clinics have achieved sufficient experience to apply the Day-5 biopsy technique.

The relative invasiveness of the embryo biopsy inherent to PGS has induced questions on its safety with regard to child development. PGS includes more extensive embryo manipulation than IVF. As it has been established that IVF and/or its underlying subfertility is associated with less favourable neurodevelopmental and cardiometabolic outcomes in offspring, it is conceivable that PGS introduces extra risk for impaired health. Yet information on development and health of PGS offspring is scarce. Most available studies addressing the sequela of embryo biopsy have evaluated the outcome of PGD and/or PGS in non-randomized study designs. We used a randomized design and reported that PGS at cleavage stage may have a minor negative effect on the neurodevelopment of IVF twins, but none on blood pressure and anthropometrics at 4 years of age.

As neurodevelopmental impairments may first emerge with increasing age, we extended the follow-up to the age of 9 years for children born to women who were randomly assigned to IVF with or without PGS. Therefore, the objective of the study was to evaluate whether PGS of cleavage stage embryo’s affects neurodevelopment, cognitive development, blood pressure or anthropometrics in IVF conceived children at the age of 9.

Materials and Methods

Study design

The PGS-trial is a multicenter trial (Academic Medical Center in Amsterdam and University Medical Center Groningen) on PGS. The trial started as a double blinded randomized controlled trial on the efficiency of PGS to improve ongoing pregnancy rates (ISRCTN76355836). The trial consisted of two groups receiving a reproductive technique (IVF or ICSI), one with PGS (PGS group) and one without PGS (control group). The trial had been prepared in the years 2001–2003; patients were not involved in the study design. Participation in the trial was not associated with increased treatment burden for patients, as the study was conducted in a blinded fashion and the treatment under evaluation took place in the laboratory, but the study showed that PGS was associated with a reduced ongoing pregnancy rate. The outcome of the trial was broadly communicated, including, but not limited to patient support groups, healthcare professionals, the health Council of the Netherlands, and the Dutch Ministry of Health, Welfare and Sport. Due to the unfavourable results of the trial, the planned cost-effect analysis was cancelled. After dissemination of the outcomes of the randomized controlled trial, the planned multistage long-term follow-up study to evaluate development and health in the children who were born, ensued. Participating families have been and are informed on the outcomes of the study and on new follow-up rounds with highly appreciated annual newsletters.

Follow-up examination

Participants are the children of subfertile women who participated in the PGS-trial. Women meeting the inclusion criteria of the PGS-trial and agreeing to participate were recruited. They received help conceiving by means of IVF and intracytoplasmic sperm injection at the Department of Reproductive Medicine of the Academic Medical Center or University Medical Center Groningen between May 2003 and November 2005 (for details of the inclusion criteria see Mastenbroek et al. 2007). In total, 408 women were included and before the first follicular aspiration randomly assigned by a computer program with a minimization procedure for age (35 through 37 years and 38 through 41 years) to a reproductive technique (IVF and intracytoplasmic sperm injection) with PGS (PGS group) or without PGS (control group). This resulted in 144 live births (Figure 1). Information on socioeconomic status, the prenatal, perinatal and neonatal period were recorded on standardized charts at the first follow-up assessment two weeks after birth (Middelburg et al., 2009). The follow-up examination at 9 years took place between March 2014 and May 2016.

The PGS-trial protocol was approved by the Medical Ethics Committee of the Academic Medical Center in Amsterdam and the follow-up study by the Medical Ethics Committee.
of the University Medical Center Groningen (Trial registration number: ISRCTN76355836). Parents provided written informed consent.

Outcome measures

Primary outcome at 9 years was the neurological optimality score, a sensitive measure of neurological condition which we assessed with the standardized, age-specific Touwen test.27 The neurological optimality score results from the sum of 64 items that fulfill criteria for optimality creating a tool sensitive for subtle neurological changes, even within the normal range.28 A decrease of 5 points is considered to be clinically relevant. The neurological assessment does not only allow to assess the neurological optimality score, but also to diagnose neurological disorders such as cerebral palsy and to assess minor neurological dysfunction. Minor neurological dysfunction is evaluated in eight domains of neurological function; each being scored as typical or deviant. Children are classified as neurologically normal (no dysfunctional domains), simple minor neurological dysfunction (1-2 domains of dysfunction, reflecting typical but non-optimal function) and complex minor neurological dysfunction (>2 domains of dysfunction, the clinically relevant form of minor neurological dysfunction).27 We considered clear neurological syndromes and complex minor neurological dysfunction as adverse neurological outcome. Reliability and validity of the assessment are satisfactory.27

Prespecified secondary outcomes were cognition, behaviour, blood pressure and anthropometrics. To assess global cognition we used the Wechsler Abbreviated Scale of Intelligence. It is a validated screening test to measure intelligence between 6 to 89 years.26 Three intelligence quotient-scores are derived: verbal intelligence; performance intelligence; and total intelligence. Specific domains of cognitive development were assessed with the Dutch version of the Neuro Psychological Assessment-II, a reliable and valid test applicable for children aged 5-12 years.29 We used three Neuro Psychological Assessment-II domains: attention and executive functions (two subtests), memory and behavioural problems.30 The number of tests with an atypical score within a domain was counted.

To assess behavioural outcome the Child Behaviour Checklist and Teacher Report Form were filled out by parents and teachers respectively. The well-validated Child Behaviour Checklist and Techer Report Form include 113 items that address emotional and behavioural problems.31 The sum of all items results in the total problem score. In addition, an internalizing problem score and externalizing problem score can be calculated.

Blood Pressure (mmHg) was measured seated using an automated blood pressure monitor (Datascope Accutorr plus, Mahwah, NJ, USA) at the non-dominant arm. Blood pressure measurement was carried out in six-fold (three times in duplo), resulting in an overall mean systolic blood pressure and diastolic blood pressure. The overall mean blood pressures were used to calculate blood pressure percentiles based on the standards of the U.S. National High Blood Pressure Education Program.32 Blood pressure percentiles take sex, height (in cm) and age (in months) into account.

Biceps, triceps, supra-iliac and sub-scapular skinfold thickness (in mm) were each measured three times, on the non-dominant side of the child, using a Servier calliper. The mean of these three measurements was used for further calculations. The mean of biceps and triceps skinfold thickness was used as a parameter of peripheral fat distribution; that of the supra-iliac and sub-scapular skinfold as a parameter of central fat distribution. The sum of the four means is an indicator of total body fat.33

The standing height (in cm) of the children was measured using a stadiometer (Seca Deutschland, Hamburg, Germany); body weight (kg) was assessed with an electronic weighing scale (Radwag, Random, Poland). Both measurements (standing height and weight) were carried out in duplicate. Information on height and weight allowed for the calculation of body mass index (kg/m²). Body mass index was classified as normal or obese according to the international classification which takes age and sex into account.34 Occipitofrontal head circumference (in cm) was measured with a non-stretchable ‘lasso’ tape.

Statistical analysis

The power calculation of the original trial was based on the number of women who were needed to detect an increase in ongoing pregnancy rates and not on the number of children needed for neurodevelopmental follow-up.25 A post-hoc power analysis showed that with our current group sizes the minimally detectable differences between the PGS group and control group was approximately 3 points on the neurological optimality score (with 80% power and α=0.05). This indicates that our sample size should be able to detect differences that are clinically relevant.

To estimate differences in child and parental characteristics between the two groups, Fisher’s exact test, Mann-Whitney-U-test and student’s t-test were used. Twins and singletons were analysed separately.

The numerical response variables length, weight, body mass index, systolic blood pressure, diastolic blood pressure, total skinfold thickness, fat distribution (central/peripheral skinfold), head circumference, neurological optimality score, intelligence and behavioural scores at 9 years were analysed with mixed effects models. The mother was incorporated as a random subject effect to model the possible correlation between the outcome variables for twins. IVF treatment with PGS was included as a fixed effect.

For the binary response variables high body mass index and neurological outcome a generalized linear mixed model with the logit link function was used. For the count variables of number of atypical tests within a neuro psychological assessment-domain the
Poisson loglinear link function was used. The estimation was performed with generalized estimating equation and the cluster variable was determined by the mother again. The robust estimator was used together with an exchangeability working correlation matrix. The type3 generalized score statistics was applied to determine the p-value for treatment effect.

All analyses were additionally corrected for twins status, IVF/Intracytoplasmic sperm injection, age at examination, time to pregnancy, maternal age, high education mother, gestational age, neonatal intensive care admission and performed with SPSS software, version 23.

Results

Participation

There were 408 women included and randomly assigned to IVF/Intracytoplasmic sperm injection with or without PGS. This resulted in 52 ongoing pregnancies in the PGS group and 74 in the control group. In the PGS group, 59 children (39 singletons; 20 twins) were born alive; in the control group 85 children (57 singletons; 14 twins) and 56 control children (38 singletons; 18 twins) participated in the study. Follow-up and attrition is summarized in the flow diagram (Fig. 1). Postnatal attrition after 9 years of follow-up was 23% in the PGS group and 32% in the control group. In general, background and short-term outcome parameters of children who were lost to follow-up at 9 years and those who were assessed were not significantly different. However, drop-out in the PGS group was associated with a higher prevalence of Caesarean section (Fisher’s exact test: P = 0.007), neonatal intensive care admission (Fisher’s exact test: P = 0.022) and being the firstborn child (Fisher’s exact test: P = 0.007). In the control group, drop-out was associated with lower maternal education (Fisher’s exact test: P = 0.038) (data not shown).

Parental and infant characteristics

Fertility parameters, obstetrical, parental and child characteristics are displayed in Table I. Gestational age was higher (39.5 weeks) in PGS pregnancies than in control pregnancies (38.7 weeks; P = 0.031); a similar difference was also present in the subgroup of singletons (PGS group: 39.7 weeks; control group: 39.1 weeks; P = 0.009). The rate of neonatal intensive care admission was higher in PGS offspring (21%) than in control offspring (5%) (P = 0.021).
### Table I. Prenatal, perinatal, neonatal and demographic characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Couples with ≥1 live birth after preimplantation genetic screening n=36</th>
<th>Control couples with ≥1 live birth n=47</th>
<th>All preimplantation genetic screening children n=43</th>
<th>All Control children n=56</th>
<th>Preimplantation genetic screening singletons n=29</th>
<th>Control singletons n=38</th>
<th>Preimplantation genetic screening twins n=14</th>
<th>Control twins n=18</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parental characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal age at conception in years, median (range)</td>
<td>37.3 (35.2-39.8)</td>
<td>37.8 (35.3-41.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education level mother (high) (^{a,b}), n (%)</td>
<td>20 (57)</td>
<td>28 (61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education level father (high) (^{a,b}), n (%)</td>
<td>20 (56)</td>
<td>22 (49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fertility parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to pregnancy in years, median (range)</td>
<td>3.4 (0.7-9.5)</td>
<td>3.8 (0.3-10)</td>
<td>3.5 (0.7-9.5)</td>
<td>3.6 (0.3-10)</td>
<td>3.9 (1.5-8.5)</td>
<td>3.9 (0.8-8.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVF without intracytoplasmic sperm injection, n (%)</td>
<td>21 (58)</td>
<td>26 (55)</td>
<td>14 (48)</td>
<td>21 (55)</td>
<td>14 (100)*</td>
<td>10 (56)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVF with intracytoplasmic sperm injection, n (%)</td>
<td>12 (33)</td>
<td>20 (43)</td>
<td>12 (41)</td>
<td>16 (42)</td>
<td>0 (0)*</td>
<td>8 (44)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-uterine insemination performed(^{d}), n (%)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naturally conceived, n (%)</td>
<td>2 (6)</td>
<td>1 (2)</td>
<td>2 (7)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preimplantation genetic screening treatment centre (UMCG), n (%)</td>
<td>21 (58)</td>
<td>25 (53)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gestational characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking during pregnancy*, n (%)</td>
<td>3 (10)</td>
<td>2 (4)</td>
<td>3 (12)</td>
<td>2 (6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol use during pregnancy*, n (%)</td>
<td>1 (3)</td>
<td>1 (2)</td>
<td>1 (4)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Birth characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caesarean section(^{e}), n (%)</td>
<td>7 (20)</td>
<td>18 (38)</td>
<td>5 (18)</td>
<td>12 (32)</td>
<td>4 (29)</td>
<td>12 (68)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational age in weeks, median (range)</td>
<td>39.5 (32-42.0)</td>
<td>38.7 (30-41.4)</td>
<td>39.7 (36.9-42.0)</td>
<td>39.1 (30-41.4)</td>
<td>37.1</td>
<td>36.7 (33.8-37.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preterm birth (&lt;37 weeks), n (%)</td>
<td>4 (11)</td>
<td>6 (13)</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>6 (43)</td>
<td>10 (56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birthweight in grams, mean (sd)</td>
<td>3266 (762)</td>
<td>3119 (681)</td>
<td>3624 (561)</td>
<td>3384 (590)</td>
<td>2526 (569)</td>
<td>2559 (507)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low birthweight, n (%)</td>
<td>9 (21)</td>
<td>10 (18)</td>
<td>2 (7)</td>
<td>1 (3)</td>
<td>7 (50)</td>
<td>9 (50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small for gestational age*, n (%)</td>
<td>3 (7)</td>
<td>2 (4)</td>
<td>2 (7)</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td>2 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apgar score 5 min &lt; 7, n (%)</td>
<td>0 (0)</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neonatal intensive care admission*, n (%)</td>
<td>9 (21)</td>
<td>3 (5)</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>8 (62)*</td>
<td>2 (11)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Child characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twins, n (%)</td>
<td>14 (33)</td>
<td>18 (32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Firstborn, n (%)</td>
<td>20 (56)</td>
<td>28 (60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>22 (51)</td>
<td>30 (54)</td>
<td>15 (52)</td>
<td>17 (45)</td>
<td>7 (50)</td>
<td>13 (72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at examination in months, median (range)</td>
<td>111 (109-124)</td>
<td>112 (109-128)</td>
<td>112 (109-124)</td>
<td>112 (109-124)</td>
<td>111 (10-121)</td>
<td>112 (110-121)</td>
<td></td>
<td></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 preimplantation genetic screening group and the control group at parent level and at child level separately for singletons and twins.

*P < 0.05

\(^{a}\) Missing data in the preimplantation genetic screening group: alcohol consumption during pregnancy n=7, Apgar score 5 min < 7 n=5, Caesarean section n=1, education level mother n=1, neonatal intensive care admission n=1, smoking during pregnancy n=4, Apgar score 5 min < 7 n=1, education level father n=2, education level mother n=1, smoking during pregnancy n=2.

\(^{b}\) University education or vocational colleges

\(^{c}\) Birthweight for gestational age is < -2 standard deviations compared with the Dutch reference population (Dutch reference tables, perinatal Registration Netherlands).

\(^{d}\) In case of poor follicle growth, treatment was converted to intra-uterine insemination.
Neurodevelopmental outcomes

Table II presents neurological, cognitive and behavioural outcomes of the PGS group and the control group. Neurological outcome of both groups was not significantly different. In the PGS group, 17 children (40%) had an adverse outcome, including one child with cerebral palsy; in the control group adverse neurological outcome occurred in 19 children (34%) with none having a neurological syndrome. Also the neurological optimality score did not differ significantly between the PGS group (mean values [95% CI]: 51.5 [49.3; 53.7]) and control group (53.1 [50.5; 55.7]). Total intelligence quotient of PGS offspring (114 [108; 120]) was not significantly different to that of controls (117 [109; 125]). The same held true for verbal intelligence, performance intelligence and specific cognitive domain scores. Behavioural outcomes of children born after PGS were also not significantly different to those of control children.

Blood pressure and anthropometrics

Table II also presents the outcomes of blood pressure and anthropometrics. Blood pressure and anthropometric outcome of the PGS group was not significantly different to that of the control group. Systolic blood pressure of the PGS group was 106 mmHg [95% CI 103;109], that of the controls was 107 [104;110]. The values of diastolic blood pressure were 66.1 mmHg [63.3;68.9] and 65.3 mmHg [62.0;68.6], respectively. Also, blood pressure percentiles did not show statistically significant differences between groups. Anthropometric values of the two groups did not differ.

Table II. Health and development of cleavage-stage preimplantation genetic screening offspring: results of the adjusted mixed-effect model analyses at 9 years of age

<table>
<thead>
<tr>
<th>Response variable</th>
<th>Preimplantation genetic screening group (n=43)</th>
<th>Control group (n=56)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurological outcome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurological optimality score, mean [CI]</td>
<td>51.5 [49.3;53.7]</td>
<td>53.1 [50.5;55.7]</td>
<td>0.193</td>
</tr>
<tr>
<td>Adverse neurological outcome*, n (%)</td>
<td>17 (40)</td>
<td>19 (34)</td>
<td>0.975</td>
</tr>
<tr>
<td>Cognitive outcome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Intelligence quotient, mean [CI]</td>
<td>114 [108;120]</td>
<td>117 [109;125]</td>
<td>0.375</td>
</tr>
<tr>
<td>Verbal Intelligence quotient, mean [CI]</td>
<td>114 [108;121]</td>
<td>114 [106;123]</td>
<td>0.937</td>
</tr>
<tr>
<td>Performance Intelligence quotient, mean [CI]</td>
<td>110 [104;117]</td>
<td>116 [108;124]</td>
<td>0.126</td>
</tr>
<tr>
<td>Anthropometrics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight (kg)*, mean [CI]</td>
<td>34.8 [31.6;38.0]</td>
<td>35.6 [32.3;38.8]</td>
<td>0.669</td>
</tr>
<tr>
<td>Standing height (cm)*, mean [CI]</td>
<td>142 [139;144]</td>
<td>142 [139;145]</td>
<td>0.866</td>
</tr>
<tr>
<td>Body Mass Index*, mean [CI]</td>
<td>17.4 [16.2;18.7]</td>
<td>17.6 [16.2;18.9]</td>
<td>0.853</td>
</tr>
<tr>
<td>Body Mass Index&gt;25kg/m², n (%)</td>
<td>11 (26)</td>
<td>10 (18)</td>
<td>0.915</td>
</tr>
<tr>
<td>Total skinfold thickness, mean [CI]</td>
<td>3.78 [2.81;4.75]</td>
<td>3.91 [2.87;4.95]</td>
<td>0.659</td>
</tr>
<tr>
<td>Central fat/peripheral fat distribution, mean [CI]</td>
<td>0.88 [0.76;0.99]</td>
<td>0.96 [0.84;1.09]</td>
<td>0.188</td>
</tr>
<tr>
<td>Head circumference (cm), mean [CI]</td>
<td>53.8 [52.5;54.4]</td>
<td>53.9 [51.2;54.7]</td>
<td>0.659</td>
</tr>
</tbody>
</table>

In the analyses we corrected for twins status, IVF/Intracytoplasmic sperm injection, age at examination, time to pregnancy, maternal age, high education mother, gestational age and neonatal intensive care admission.

* Adverse neurological outcome consisted of clear neurological syndromes (one child with cerebral palsy in the PGS group) and complex minor neurological dysfunction.

The blood pressure percentiles take sex, height and age in months into account.
Discussion

This multicentre follow-up study of a randomized controlled trial indicates that PGS in cleavage stage embryos is not associated with adverse effects on neurological, cognitive and behavioural development, blood pressure or anthropometrics in offspring at 9 years.

At the age of 2 years, we reported that PGS offspring had a lower neurological optimality score than controls.\textsuperscript{35} At the age of 4 years, only PGS twins had a lower neurological optimality score.\textsuperscript{23} The current findings indicate that the negative association between PGS and neurological condition does not persist until 9 years. Two explanations may be offered for the disappearance of the association. First, the plastic changes of the developing brain may result in recovery of dysfunction.\textsuperscript{26} Second, it is possible that the effects of PGS are so minimal that they could not be detected with current group sizes, even with the use of the sensitive neurological optimality score. Our findings are in line with the only study that has investigated the safety of embryo biopsy (PGD) for cognitive and psychomotor development beyond the age of 5–6 years. However, that study had an observational design and did not address the effect of PGS on top of IVF (Winter et al., 2014). It should be noted, however, that the prevalence of adverse neurological outcome in our study was considerably higher than that of the general population (36 versus 5–7%).\textsuperscript{27,37}

This underlines the previously mentioned less favourable neurodevelopmental outcome of children born after IVF. The large majority of children with an adverse neurological outcome showed the complex form of minor neurological dysfunction. This non-optimal neurological condition is associated with an increased risk of developmental disorders, such as developmental coordination disorder, attention deficit hyperactivity disorder, autism spectrum disorders, coordination problems and specific learning problems.\textsuperscript{37–40}

Further studies are needed to unravel the underlying mechanisms.

Blood pressure and anthropometric outcome is in line with our findings at the age of 4 and with those of others in children born after PGD and/or PGS at 2 years.\textsuperscript{26,29} Our study suggests that PGS does not have a long-term effect on blood pressure, i.e. at the age of 9 years. Nevertheless, it is noteworthy that the mean BP percentiles in both groups are slightly above the 60th percentile. In line with the current literature, this suggests that IVF offspring are vulnerable to an increased BP.\textsuperscript{15,18,19} Underlying mechanisms are still not clear. Therefore more research is necessary.

A strength of the study is the long-term follow-up of children born after IVF, where couples were randomly assigned to IVF with or without PGS. This design resulted in two groups with comparable background variables, so that the comparison of the PGS group and control group primary reflects the effect of PGS. Another strength is that we only studied children born following PGS. Most studies evaluating embryo biopsy included children born following PGD and PGS. Yet, where the technical procedures of PGD and PGS are similar, the indication for both techniques is not. PGD is used to avoid transmission of hereditary diseases in otherwise fertile couples, whereas PGS is used to increase the effectiveness of treatment in subfertile couples undergoing IVF.\textsuperscript{41} As subfertility is associated with obstetrical problems, perinatal adversities and a less optimal neurological development, PGS offspring could be more at risk for adverse health outcome than PGD offspring.\textsuperscript{17,42–44}

Our study had limitations. The power calculation of the original study was not based on the number of children needed for the follow-up study. PGS reduced the ongoing pregnancy rates, which further reduced sample size for follow-up. However, the post-hoc power analysis indicated that the current study should have been able to detect differences that are clinically relevant. Both singletons and twins participated in the study. To avoid a potentially confounding effect of twin status on outcome, outcome appropriate mixed effects models with additional adjustment for twin status were applied.\textsuperscript{44} We realize that the generalizability of our findings is hampered due to the relatively high intelligence quotient of the participating children, possibly related to the well-educated parents and the selective drop-out of low educated parents.\textsuperscript{45}

Our study evaluated embryo biopsy in the form of PGS at cleavage stage (Day-3 embryo biopsy), while in general the current practice is PGD at blastocyst stage (Day-5 embryo biopsy).\textsuperscript{46} Yet, we consider our long-term follow-up data as clinically relevant. First, the majority of embryos biopsied prior to 2012, received their biopsy at Day-3 cleavage-stage embryos.\textsuperscript{12} For the healthcare professionals in charge of the counselling of the couples who received PGS with Day-3 embryo biopsy, it is important to know whether PGS offspring are at increased risk of non-optimal health and development. Second, Day-3 biopsy is still performed in IVF clinics.\textsuperscript{13} The clinical community would certainly benefit from a novel well-powered randomized controlled trial evaluating the effect of PGS with Day-5 embryo biopsy in combination with a long-term follow-up of offspring health and development.

In conclusion, in this follow-up study of a randomized controlled trial, we found no evidence of adverse consequences of Day-3 cleavage-stage PGS on neurodevelopment and health of offspring at 9 years. These findings are reassuring for couples considering PGS and for parents of PGS offspring. However, both groups of IVF offspring showed high prevalences of the clinically relevant form of minor neurological dysfunction, which is a point of concern for the IVF community. In addition, our study confirms findings of others that IVF offspring maybe at risk of an unfavourable cardiovascular outcome. These findings are alarming and highlight the importance of research on the underlying mechanisms of unfavourable neurodevelopmental and cardiovascular outcomes of IVF offspring.

Acknowledgements

We are greatly thankful to the parents and children who participated in the study, and to Anneke Kracht and Linze Dijkstra for their technical assistance.
References


