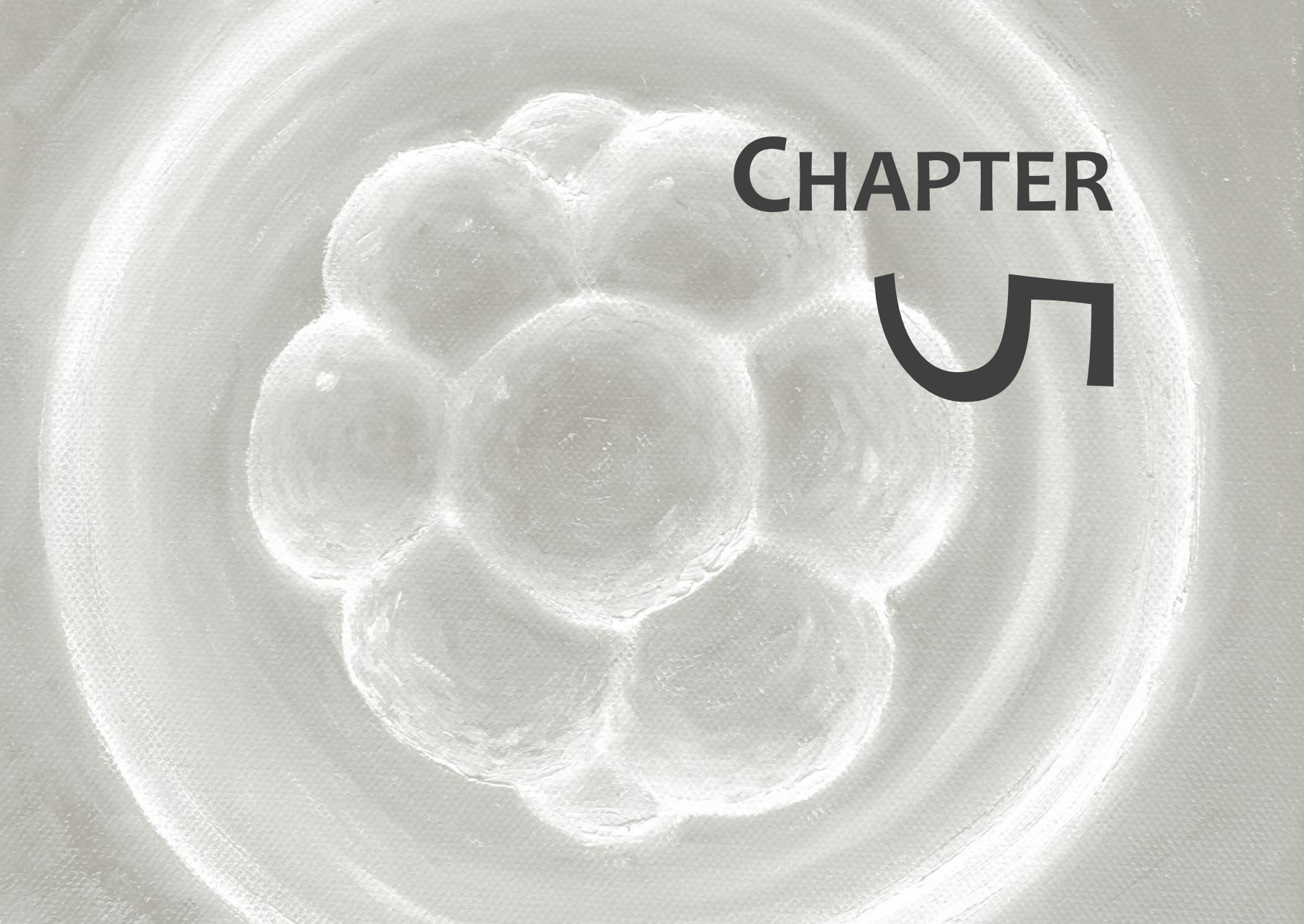

Chapter 5

DAY 3 EMBRYO SELECTION BY METABOLOMIC PROFILING OF CULTURE MEDIUM WITH NEAR-INFRARED SPECTROSCOPY AS AN ADJUNCT TO MORPHOLOGY: A RANDOMIZED CONTROLLED TRIAL

C.G. Vergouw
D.C. Kieslinger
E.H. Kostelijk
L.L. Botros
R. Schats
P.G.A. Hompes
D. Sakkas
C.B. Lambalk

Human Reproduction 2012; 27: 2304-2311

ABSTRACT

Study question: Is the selection of a single day 3 embryo by metabolomic profiling of culture medium with near-infrared (NIR) spectroscopy as an adjunct to morphology able to improve live birth rates in IVF, compared to embryo selection by morphology alone?

Summary answer: The live birth rate after embryo selection by NIR spectroscopy and morphology is not significantly different compared to the live birth rate after embryos were selected by morphology alone.

What is known already: The elevated incidence of pregnancy and neonatal problems associated with a high-twinning rate after IVF can only be successfully reduced by the transfer of one embryo. Current embryo assessment methods are unable to accurately predict the reproductive potential of an individual embryo. Today, a number of techniques are said to be more accurate at selecting the best embryo. One of these new technologies is metabolomic profiling of spent embryo culture media with the use of NIR spectroscopy.

Study design, size, duration: A double blind, randomized controlled trial was conducted between 2009 and 2011, and included 417 couples undergoing IVF with a single embryo transfer (SET). Randomization was performed centrally just before Ovum Pick-Up (OPU), using a computerized randomization program. Both patient and physician were unaware of treatment allocation. To ensure blinding, the allocations were placed in consecutively numbered, opaque envelopes. Patients were randomized (1:1) into either the control group (embryo selection by morphology only) or the treatment group (embryo selection by morphology plus NIR spectroscopy of embryo culture medium).

Participants/materials, setting, methods: At OPU, 208 patients were randomized to the morphology only group and 209 patients were randomized to the morphology plus viability score group. On day 3, 163 patients in the control group and 146 patients in the treatment group met the inclusion criteria. The study was conducted in an academic hospital with IVF laboratory and three non-academic hospitals.

Main results and the role of chance: Patient demographics and baseline characteristics were distributed equally over the two groups, except for embryo fragmentation, which was significantly higher in the treatment group. In the intention-to-treat analysis, the live birth rates were 31.7 and 26.8 % for the control group and the treatment group, respectively (relative risk 0.84; 95% confidence interval 0.63- 1.14, P=0.27). In the per protocol analysis, the live birth rates were 31.3 and 29.5% for the control group and the treatment group, respectively (relative risk 0.94; 95% confidence interval 0.67-1.32,

P=0.73). For the treatment group, the embryological technician's independent choice (by morphology) of which embryo to transfer was recorded 138 times. In 75.4% (104 of 138) of the transfers, the embryo with the best morphology did not have the highest viability score. The live birth rate of these 104 transferred embryos was 30.8%.

Limitations, reasons for caution: A possible limitation of our study is the pre-selection of all embryos by morphology and dividing the cohort of available embryos into two groups: good quality embryos and poor quality embryos. As a consequence, we have probably selected for a better prognosis patient group.

Wider implications of the findings: To avoid the use of incompetent embryo selection tools at the expense of the patient, evidence-based proof of clinical usefulness is essential before the implementation of new diagnostic tools in IVF laboratories.

Trial registration number: Dutch Trial Register, registry number NTR1178.

INTRODUCTION

IVF treatment produces a high risk of twin and high order pregnancy, which is associated with an elevated incidence of pregnancy and neonatal problems compared to singletons[1,2]. This complication is mainly caused by the transfer of multiple embryos in order to maximize pregnancy chances per IVF cycle. An increased use of elective single embryo transfer (eSET) is the only means to reduce multiple pregnancy rates[1,3,4]. Because only one embryo is transferred, the selection of the embryo with an optimum implantation potential is vital. Currently, routine embryo selection is mainly focused on morphological criteria and cleavage rate during embryo development, assessed at several predetermined microscope evaluations. This is a highly subjective method[5,6] and its relation to pregnancy outcome is uncertain[7]. The development of an accurate method to assess the embryo viability which will lead to equal or improved pregnancy rates while decreasing the incidence of twinning is currently one of the main challenges for embryologists[8]. Non-invasive embryo selection methods, such as the observation of the embryo itself or the analysis of its environment (culture medium), are preferred.

During the last decades, new non-invasive embryo selection methods that analyse culture medium constituents have been studied[9]. An embryo uses nutrients and generates metabolites, which means that the embryo changes the composition of its culture medium. These changes reflect the activity and efficiency of embryo metabolism, which may be used as a predictor of embryo viability: viable embryos modify their culture medium differently compared to non-viable embryos. One of these methods, analysing the changes in culture medium made by the embryo, is a screening technology using near-infrared (NIR) spectroscopy. A NIR spectrometer simultaneously measures the quantitative absorption of key functional groups such as -CH, -OH and -NH in a fluid and generates a correlating spectrum. These key functional groups are associated with culture medium constituents such as glucose, pyruvate, albumin, glutamate and lactate. Several studies have reported the use of NIR spectroscopy to measure small alterations of the chemical composition of culture media made by the embryo, after the principle of using these types of technology for this purpose was suggested by the group of Goodacre[10]. Embryos with positive and negative pregnancy outcomes showed distinct differences in culture medium composition. Spectral profiles of spent embryo culture medium were compared between embryos that resulted in positive and negative pregnancy outcomes. The regions within the NIR spectra between 920 and 1675 nm that discriminated between implanted and non-implanted embryos were quantified in a multivariate algorithm and subsequently expressed as a viability score[11-18].

The proof of principle studies that used NIR spectroscopy to generate a viability score consistently showed significant higher mean viability scores for embryos that resulted in an ongoing pregnancy and/or live birth, compared to embryos with a negative pregnancy outcome, in fresh[11-18] and frozen-thawed[19] embryo transfer cycles. This was seen regardless of day of transfer and was independent of embryo morphology. Furthermore, the morphology and viability score combined proved to be more accurate in predicting pregnancy results[17] and the viability score was a better predictor of pregnancy outcome than blastocyst morphology[18].

These findings using NIR spectroscopy are promising. Nevertheless, the studies were mainly retrospective in design and spent embryo culture medium had to be snap frozen and sent to a single laboratory for analysis. Therefore, whether metabolomic profiling using NIR spectroscopy as an adjunct to morphology will prove a robust test when performed at IVF sites is not yet known[8], although a premature interim analysis of a prospective study evaluating day 2 or day 5 embryo culture medium with ongoing pregnancy as primary outcome could not show a benefit[20]. We here report on a complete randomized controlled trial (RCT) that evaluated the use of metabolomic profiling in addition to standard morphological selection on cultured day 3 embryos with live birth rate as primary end-point.

MATERIALS AND METHODS

STUDY DESIGN

The study was performed as a double blind RCT to investigate if the selection of a single day 3 embryo by metabolomic profiling of culture medium with NIR spectroscopy as an adjunct to morphology is able to improve live birth rates in IVF, compared to embryo selection by morphology alone. Inclusion started in 2009 and ended in 2011. Patients scheduled for IVF or ICSI with a SET at the IVF center of VU University medical center in Amsterdam, or one of its affiliated clinics, were asked to participate in the study. The stimulation phase and oocyte retrieval were performed at VU University medical center or one of its affiliated clinics, the laboratory phase and embryo transfer were exclusively performed at VU University medical center.

All participating patients signed an informed consent. The study protocol was approved by the Central Committee on Research involving Human Subjects (CCMO) and by the institutional review board of all participating clinics. The trial was registered as an RCT in the Dutch Trial Register, with registry number NTR1178.

STIMULATION PROTOCOL

Women aged 38 years or younger or with previous good response to IVF or ICSI treatment underwent controlled ovarian hyper stimulation with a long GnRH agonist (triptorelin [Decapeptyl®; Ferring, Denmark] or leuprorelin [Lucrin®; Abbott, USA]) or GnRH antagonist (nafarelin [Synarel®; Pfizer, USA] or cetrorelix [Cetrotide®; Merck Serono, Germany]) protocol. In women older than 38 years or with a previous poor response, a short GnRH agonist or GnRH antagonist protocol was applied. Ovarian stimulation was achieved with recombinant FSH (Gonal-F®; Merck Serono, Germany or Puregon®; MSD, USA), human menopausal gonadotropin (Menopur®; Ferring, Denmark) or urofollitropin (Fostimon®; IBSA Farmaceutici, Italy). The ovarian response was monitored by vaginal ultrasonography and serum estradiol determinations. Human chorionic gonadotropin (Pregnyl®; Organon, the Netherlands) 10,000 IU s.c. was given 36 h before ultrasonographic-guided oocyte retrieval, when there was at least one follicle ≥18 mm and three or more follicles ≥16 mm. The luteal phase was supported by progesterone intravaginally (3X200 mg daily; Utrogestan®; Besins Health Care, Belgium) from the day of oocyte retrieval.

IVF/ICSI PROCEDURE

On the day of oocyte retrieval (day 0), IVF and ICSI were performed according to the laboratory's routine insemination procedures. IVF oocytes were placed in fertilization medium (Sage®; Quinn's advantage protein plus fertilization medium, Cooper Surgical, USA) until fertilization was checked 16-18 h after insemination. At that time, the IVF zygotes were transferred into 25-µl pre-equilibrated medium drops of cleavage medium (Sage®; Quinn's advantage protein plus cleavage medium, Cooper Surgical, USA). ICSI oocytes were placed directly into 25-µl pre-equilibrated cleavage medium drops after injection. Embryos were cultured individually under oil in incubators at 37°C, under 5% CO_2 and atmospheric O_2 concentration. Embryos were morphologically assessed 25-27 h after insemination (early cleavage score) and on day 2 and day 3 between 8.00 and 9.00 h in the morning. In the morning of day 3, embryos were transferred into a new culture dish with blastocyst medium (Sage®; Quinn's advantage protein plus blastocyst medium, Cooper Surgical, USA). Embryo transfer was performed on day 3, early in the afternoon.

Good quality embryos that were not transferred were frozen on day 4. All remaining (poor quality) embryos were checked for blastocyst formation on day 6.

EMBRYO SELECTION FOR TRANSFER

Each embryo was morphologically assessed by combining the number and regularity of blastomeres and the degree of fragmentation. On day 3, embryos were divided into two groups: good quality embryos (embryos with at least six cells and ≤20% fragmentation) or poor quality embryos (embryos with < six cells and/or embryos with >20% fragmentation). When there were at least two embryos of similar best quality, the study was continued. When there was only one good quality embryo available, patients were excluded prior to transfer. In the treatment group, a viability score for each embryo was generated using a prototype NIR spectroscopy instrument and pre-determined algorithm. All embryos in the treatment group were analysed, but only the good quality embryos with at least six blastomeres and ≤20% fragmentation were considered for embryo transfer. If there was no good quality embryo available for transfer, the same procedure was followed as within a group of good quality embryos. The embryo with the highest viability score within the group of embryos of similar best quality was transferred in the treatment group. The embryo with the best morphology was transferred in the control group.

The morphological choice of which embryo to transfer was based on the laboratory's routine procedures. Overall, an embryo with 0% fragmentation was favoured over an embryo with 1-20% fragmentation and >20% fragmentation respectively. Furthermore, an eight cell embryo was favoured over a seven cell, six- or nine cell, 10-cell, compaction stage, five cell and four cell embryo respectively. Therefore, a top quality embryo is defined as an eight cell embryo without fragmentation. When there were two or more exactly the same embryos on day 3, the day 2 and early cleavage score were used for embryo selection.

NIR SPECTROSCOPY METHOD

A prototype NIR spectroscopy system (Molecular Biometrics Inc., Norwood MA, USA) intended for IVF use was used to analyse spent embryo culture media on day 3 for the treatment group. The spectrometer was checked daily for any wavelength calibration offsets, electronic noise, and lamp drift, using a qualified standard, and dark cell measurements, respectively. For each cohort measurement, a dark measurement was also measured for mathematical dark subtraction from resulting embryo profiles. For each embryo, an instrument-specific sample cell was filled with 10-µl cleavage medium from each separate media droplet in which the embryo was cultured. Sample cells were placed for at least 4 min in a temperature stabilizer at 24.3°C, after which it was transferred into the NIR spectrometer. All samples were analysed three times to rule out any measurement failures. The measurement was repeated with the control media

to account for any variations between culture conditions between embryos. Control medium was of the same batch and was kept in an incubator in a separate tube under the same conditions as the embryos. After measuring the control media, the NIR spectrometer produced a viability score using a multivariate algorithm developed using NIR spectra with known pregnancy outcome of transferred day 3 embryos. This algorithm was developed using proprietary methodology of Molecular Biometrics Inc. entailing a modification of previous publications[13,16].

END-POINTS, INCLUSION CRITERIA, EXCLUSION CRITERIA AND RANDOMIZATION

The primary end-point of the study was live birth rate. The secondary end-point was ongoing pregnancy rate, defined as a viable intrauterine pregnancy at 10 weeks after embryo transfer. Randomization was performed centrally just before Ovum Pick-Up (OPU), using a computerized randomization program. Both patient and physician were unaware of treatment allocation. To ensure blinding, the allocations were placed in consecutively numbered, opaque envelopes. Patients were randomized (1:1) into either the control group (embryo selection by morphology only), or the treatment group (embryo selection by morphology plus NIR spectroscopy). Main inclusion criterion prior to SET was the presence of two or more similar best quality embryos. Main exclusion criteria were the transfer of more than one embryo and patients were only allowed to participate once. Only cycles with ejaculated sperm were included.

SAMPLE SIZE CALCULATION

A minimal increase of 15% in live birth rate was considered to be clinically relevant. In order to detect a 15% increase, with a power of 80% at the 5% level of significance, 153 transfers were needed in each group. A withdrawal percentage of 20% of patients that did not meet the inclusion criteria at time of embryo transfer was included.

STATISTICS

We performed an intention-to-treat analysis (ITT) with data from all randomized patients and a per protocol (PP) analysis, with data from patients who completed the intervention. We calculated ongoing pregnancy rates and live birth rates per OPU for the control and treatment group and the corresponding relative risk with 95% confidence intervals. Chi-square tests were used to test for significance.

The following (maternal) baseline characteristics were compared between the treatment group and the control group: age, body mass index (BMI), percentage of primary infertility, duration of infertility, medical cause of infertility, number of previous IVF attempts, type of pituitary regulation, total dosage of gonadotropins administered,

fertilization method, number of oocytes retrieved, number of fertilized oocytes, number of good quality embryos available at transfer, number of cells of the embryo that was transferred, degree of fragmentation of the embryo that was transferred and the number of embryos that were cryopreserved. Student's t-tests and chi-square tests were used to compare mean values of these patient demographics.
In the treatment group, Student's t-tests were used to compare the mean viability scores between implanted and non-implanted embryos. Additionally, a descriptive analysis was conducted as per the embryological technician's independent morphological choice of which embryo to transfer. This was done by asking the embryological technician to choose the best embryo based on morphology. The embryological technician was unaware of the viability scores of the embryos.

RESULTS

A total of 417 patients were randomly assigned to treatment (Figure 1); 208 patients were assigned to the control group (embryo selection by morphology only) and 209 patients were assigned to the treatment group (embryo selection by morphology plus NIR spectroscopy). There were 163 patients in the control group and 146 patients in the treatment group who completed intervention and had an embryo transfer. The main reason of exclusion in both groups was the absence of at least two embryos of similar best morphology. The exclusion rate was higher in the treatment group due to NIR technology problems, which means that no viability score could be generated. For these 18 patients, embryo selection was performed by morphology alone.
For both ITT and PP analyses, patient baseline characteristics were equally distributed (Table 1), except for embryo fragmentation of the embryo selected for transfer, which was significantly higher in the treatment group. The number of top quality embryos transferred was significantly higher in the morphology only group.
A viability score was calculated for 1071 embryos in the treatment group: 715 embryos were of good quality (at least six cells and ≤20% fragmentation) at the time of embryo transfer. The highest viability score of these 715 good quality embryos was 1.60 and the lowest viability score was -0.77. The viability scores of the 356 poor quality embryos (< six cells and/or >20% fragmentation) ranged from -0.54 to 1.32. Some patients (n=34) had embryos of poor morphological quality with higher viability scores than their sibling embryos of good quality. One patient had two embryos with the same highest viability score, therefore embryo selection for this patient was based on morphology alone.

Table 1. Baseline characteristics and characteristics of treatment cycles (intention-to-treat [ITT] and per protocol [PP] analysis).

	ITT		PP	
	Morphology only (n=208)	Morphology + viability score (n=209)	Morphology only (n=163)	Morphology + viability score (n=146)
Age (years)	34.0 ± 4.4	34.5 ± 4.1	34.0 ± 4.5	34.6 ± 4.1
BMI (kg/m2)	24.67 ± 4.52 n=205	24.35 ± 4.53	24.91 ± 4.56 n=162	24.15 ± 4.78 n=145
Primary infertility	115 (55.3)	126 (60.3)	92 (56.4)	89 (61.0)
Duration of infertility (years)	3.29 ± 2.07	3.14 ± 2.02	3.33 ± 1.98	3.17 ± 2.39
Medical cause of infertility				
male	98 (47.1)	104 (49.8)	82 (50.3)	76 (52.1)
tubal	26 (12.5)	36 (17.2)	20 (12.3)	26 (17.8)
unexplained	42 (20.2)	33 (15.8)	31 (19.0)	21 (14.4)
endometriosis	17 (8.2)	18 (8.6)	10 (6.1)	13 (8.9)
other	25 (12.0)	18 (8.6)	20 (12.3)	10 (6.8)
Number of previous IVF attempts	0.32 ± 0.64	0.31 ± 0.63	0.31 ± 0.62	0.34 ± 0.67
Type of pituitary regulation				
Decapeptyl®	155 (74.5)	143 (68.4)	124 (76.1)	104 (71.2)
Lucrin®	50 (24.0)	60 (28.7)	36 (22.1)	38 (26.0)
Cetrotide®	2 (1.0)	4 (1.9)	2 (1.2)	2 (1.4)
Synarel®	1 (0.5)	1 (0.5)	1 (0.6)	1 (0.7)
none	0	1 (0.5)	0	1 (0.7)
Total dosage of gonadotropins administered	2080 ± 992	2185 ± 1001	2127 ± 889	2022 ± 1024
Gonal-F®	2354 ± 900	2245 ± 767	2389 ± 929	2221 ± 731
Puregon®	1994 ± 861	1832 ± 727	2052 ± 845	1828 ± 793
Menopur®	1896 ± 641	2405 ± 1295	2077 ± 978	2360 ± 1168
Fostimon®	n/a	n/a	6400	n/a
Fertilization method				
IVF	105 (50.5)	111 (53.1)	75 (46.0)	75 (51.4)
ICSI	103 (49.5)	98 (46.9)	88 (54.0)	71 (48.6)
Oocytes retrieved	11.0 ± 6.4	11.5 ± 7.1	12.0 ± 6.1	12.4 ± 6.9
Fertilized oocytes (2PN)	6.5 ± 4.1	6.4 ± 4.6	7.4 ± 3.9	7.4 ± 4.5
Good quality embryos available at transfer	3.59 ± 2.71	3.97 ± 3.57	4.32 ± 2.59	4.91 ± 3.36
Number of cells of ET embryo				
no transfer	7 (3.4)	10 (4.8)	n/a	n/a
4	4 (1.9)	7 (3.3)	1 (0.6)	1 (0.7)
5	2 (0.96)	5 (2.4)	1 (0.6)	4 (2.7)
6	9 (4.3)	30 (14.4)	3 (1.8)	21 (14.4)
7	30 (14.4)	41 (19.6)	18 (11.0)	33 (22.6)
8	147 (70.7)	87 (41.6)	133 (81.6)	64 (43.8)
9	6 (2.9)	12 (5.7)	5 (3.1)	8 (5.5)
10	1 (0.48)	2 (0.96)	1 (0.6)	1 (0.7)
>10	2 (0.96)	15 (7.2)	1 (0.6)	14 (9.6)
Degree of fragmentation of ET embryo				
1 (0%)	116 (57.7)[a]	74 (37.2)[a]	103 (63.2)[b]	56 (38.4)[b]
2 (1-20%)	82 (40.8)[a]	119 (59.8)[a]	58 (35.6)[b]	89 (61.0)[b]
3 (>20%)	3 (1.5)[a]	6 (3.0)[a]	2 (1.2)[b]	1 (0.6)[b]
Top quality embryos transferred	97 (46.6)[c]	41 (19.6)[c]	91 (55.8)[c]	30 (20.5)[c]
Embryos cryopreserved	4.40 ± 2.69	4.96 ± 3.66	4.55 ± 2.70	4.94 ± 3.59

For continuous variables mean ± SD is presented; for categorical variables n (%) is presented; n/a = not applicable; ET = embryo transfer.

Data with same superscript differ significantly.

[a, b] P<0.001

[c] P<0.0001

The mean viability score (± SD) of embryos resulting in a live birth (0.81 ± 0.19) was not significantly different from the mean viability score of embryos that did not result in a live birth (0.79 ± 0.18), P=0.64.

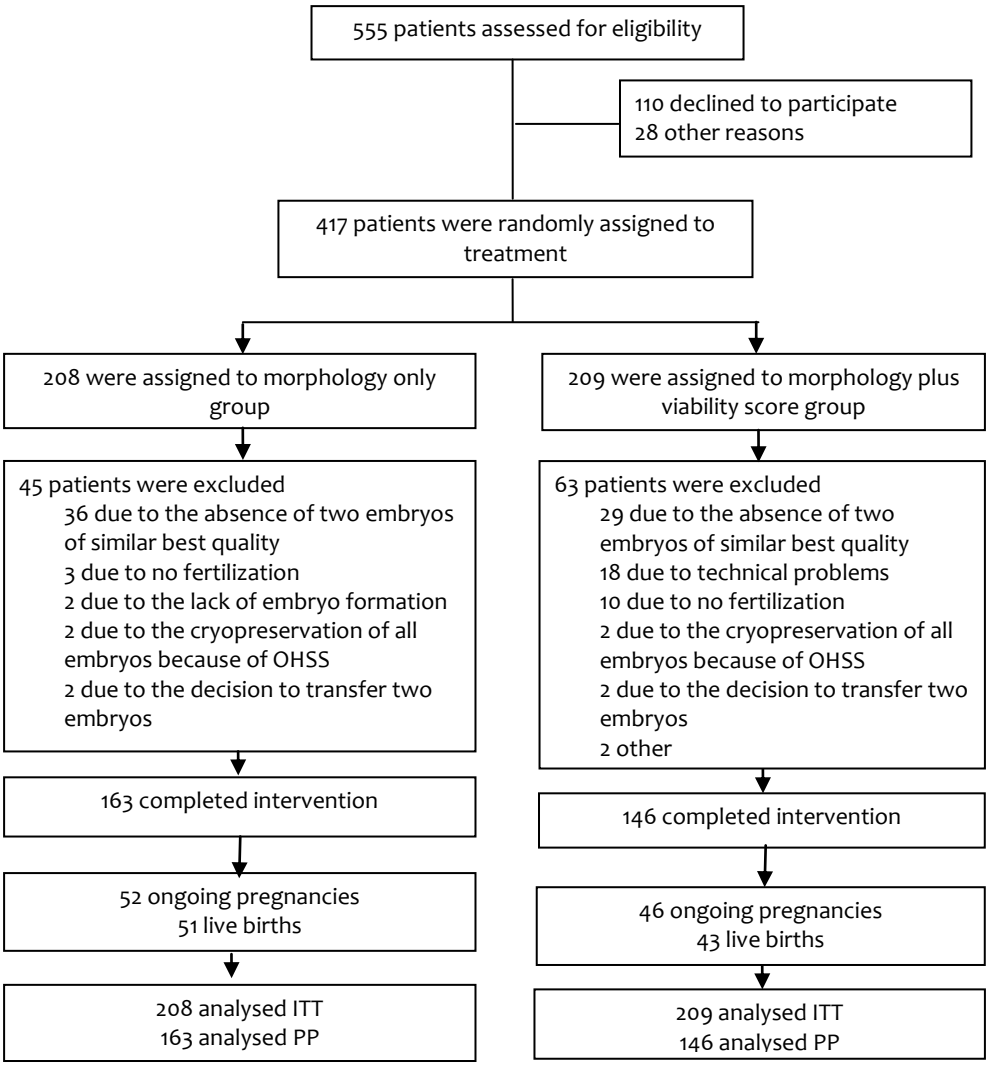

Figure 1. Flowchart of assignment, treatment and analysis of patients.

The ongoing pregnancy and live birth rates for the two study groups in the ITT and PP analyses are shown in Table 2. In the ITT analysis, there was no significant difference between the ongoing pregnancy rates of the treatment group (28.2%; 59 of 209) and the control group (32.2%; 67 of 208); relative risk 0.87, 95% confidence interval 0.65-1.17 (P=0.38). In the treatment group, 56 babies were born alive. In the control group 66 babies were born alive. The live birth rates did not differ significantly between the control and treatment groups (26.8% treatment group; 31.7% control group; P=0.27; relative risk 0.84, 95% confidence interval 0.63-1.14).

In the PP analysis, the ongoing pregnancy rates did not differ significantly between the two study groups: 31.9% (52 of 163) in the control group and 31.5% (46 of 146) in the treatment group; relative risk 0.99, 95% confidence interval 0.71-1.37 (P=0.94). In the treatment group, 43 babies were born alive and in the control group, 51 babies were born alive. The live birth rates did not differ significantly between the control and treatment groups (29.5% treatment group; 31.3% control group; P=0.73; relative risk 0.94, 95% confidence interval 0.67-1.32).

For the treatment group, the embryological technician's independent choice (by morphology) of which embryo to transfer was recorded 138 times. In 75.4% (104 of 138) of the transfers, the embryo with the best morphology did not have the highest viability score. There was no significant difference between the live birth rates of these 104 transferred embryos (30.8%; 32 of 104) and the transferred embryos in the control group (31.9%; 52 of 163); relative risk 0.96, 95% confidence interval 0.67-1.39 (P=0.85).

Table 2. Ongoing pregnancy and live birth rates (intention-to-treat [ITT] and per protocol [PP] analysis).

	ITT			
	Morphology only (n=208)	Morphology + viability score (n=209)	Relative Risk (95% CI)	P-value
Ongoing pregnancy - n (%)	67 (32.2)	59 (28.2)	0.87 (0.65-1.17)	0.38
Live birth - n (%)	66 (31.7)	56 (26.8)	0.84 (0.63-1.14)	0.27
	PP			
	Morphology only (n=163)	Morphology + viability score (n=146)	Relative Risk (95% CI)	P-value
Ongoing pregnancy - n (%)	52 (31.9)	46 (31.5)	0.99 (0.71-1.37)	0.94
Live birth - n (%)	51 (31.3)	43 (29.5)	0.94 (0.67-1.32)	0.73

DISCUSSION

This double blind RCT showed that day 3 embryo selection by metabolomic profiling of culture medium with NIR spectroscopy as an addition to morphology was not able to improve ongoing pregnancy and live birth rates compared to embryo selection by morphology alone.

Our findings are consistent with data from an interim analysis of another RCT with a similar study design but for day 2 and day 5 SET[20]. In this interim analysis, no significant difference was found in the combined day 2 and day 5 data with regard to the ongoing pregnancy rates, when an embryo was selected by morphology plus the viability score compared to embryo selection by morphology alone. This interim analysis, however, was performed earlier than planned: only 37% of the inclusions were met, while the interim analysis was planned at 50% of the inclusions. Therefore, their conclusions are debatable and their number of included patients who completed intervention per day of transfer (n=170 for day 2 and n=157 for day 5 embryo transfer) was much lower than our included number of patients for day 3 embryo transfer (n=309).

We could not reproduce the findings of previous retrospective studies using the same viability assessment method in which higher viability scores were consistently associated with higher pregnancy outcomes[11-18]. Beside the similar pregnancy and live birth rates between the treatment group and control group in our study, no significant difference was found in the treatment group between the mean viability score of embryos resulting in a live birth and the mean viability score of embryos that did not result in a live birth. This lack of reproducibility might be due to different conditions used in the retrospective trials: all culture media were snap frozen and sent to one laboratory where all analyses were done. The viability score test is possibly not as robust when performed at IVF sites. This is probably caused by the susceptibility of the used algorithms to noise which lowers the precision and repeatability of the viability score[20,21]. Additionally, inconsistencies between retrospective trials and RCTs are not unusual[20].

We excluded 26% of the randomized patients because they did not meet the inclusion criteria (SET and at least two embryos of similar best morphology available prior to transfer). This was higher than the anticipated 20% and was therefore compensated at the end of the study by including more patients. The main reason of exclusion in both groups was the absence of at least two embryos of similar best morphology. The exclusion rate was higher in the treatment group due to technical NIR problems, resulting in a failure to generate a viability score. This included the mixing of several drops of embryo culture media, the appearance of air bubbles in the sample cell and

malfunction of the NIR spectrometer. Failed NIR analysis also led to embryo selection by morphology alone.

The embryological technician was asked to identify the morphologically best embryo in the treatment group, to analyse the percentage of patients where another embryo would have been transferred when selection was based on morphology plus the viability score instead of morphology alone. The embryological technician was unaware of the viability scores of the embryos. In 75.4% (104 of 138) of the transfers, the embryo with the best morphology did not have the highest viability score. The live birth rate in this group of 104 embryos (30.8%) was not significantly different from the live birth rate in the control group (31.3%). This strongly suggests that within a group of good quality embryos, there is more than one embryo able to develop into an ongoing pregnancy. Therefore, to avoid multiple pregnancies, we stress that it is better to transfer one embryo and cryopreserve the others when several good quality embryos are available.

Although the cohort of good quality embryos was the same between the two study groups, the morphology of the transferred embryos was lower in the morphology plus viability score group (significantly less top quality embryos, i.e. eight cell embryos without fragmentation on day 3 were transferred). Therefore, it was expected that the live birth rate could be lower in the morphology plus viability score group, but it was similar to the morphology only group. We indicate that the NIR spectrometer did indeed distinguish between viable and non-viable embryos, only not better, or more accurately, than embryo selection by morphology alone. Therefore, in its current state, the NIR technique has no clinical benefit and the need for a tool that identifies the most viable embryo within a cohort remains. Hardarson et al.[20] likewise showed no significantly altered pregnancy rates although embryos in the morphology plus viability score group were morphologically of lower quality. The significantly higher percentage of fragmentation in embryos transferred in the treatment group was caused by the difference in embryo selection method between the two groups. In the control group, embryos with the least fragmentation, preferably 0%, were transferred. In the treatment group, embryos were not selected by lowest degree of fragmentation; they could have up to 20% fragmentation.

A possible limitation of our study is the pre-selection of all embryos by morphology and dividing the cohort of available embryos into two groups: good quality embryos and poor quality embryos. As a consequence, we have probably selected for a better prognosis patient group. The current study design as such was chosen deliberately. For us it was unethical to transfer an embryo of poor morphological quality when a morphological good quality embryo was available as well. After all, there is no firm evidence yet for a possible benefit of the viability score. There were 34 patients who

had embryos of poor morphological quality with a higher viability score than their sibling embryos of good quality. According to protocol, these poor morphological embryos were not considered for transfer. The embryo that indeed was transferred had the highest viability score within the patient's cohort of embryos with at least six cells and ≤20% fragmentation. The small number of patients (nine) in our study whose embryos were all of poor morphological quality were not pregnant and the remaining embryos showed no blastocyst formation on day 6 of the non-transferred embryos.

There are now several examples of embryo selection tools that have shown very promising results in retrospective trials, but appeared not to add value when they were tested in a randomized trial. The current study and the study by Hardarson et al.[20] showed no significant differences in pregnancy outcomes between embryo selection with a new, non-invasive embryo selection tool (viability score) added to morphology and embryo selection by morphology alone; although previous retrospective studies showed a promise of better pregnancy outcomes. Likewise, looking, for example, at preimplantation genetic screening, the promise of better pregnancy results was undermined by several RCTs[22]. As the search for new tools to improve the efficacy of embryo selection will continue, we emphasize that it is very important to perform RCTs to study the effectiveness of a new tool in daily IVF practice. By doing so, a better judgement of the direct relevance to biological outcome (i.e. live birth) can be made.

In conclusion, looking at our results and the results from the study by Hardarson et al.[20], there is at present no evidence that NIR spectroscopy of spent embryo culture media in its current form can be used in daily practice to improve the live birth rates after SET. Furthermore, evidence-based proof of clinical usefulness is essential before the implementation of new diagnostic tools to assess the embryo viability.

ACKNOWLEDGEMENTS

The authors would like to thank all employees of the IVF center of VU University medical center for their enthusiastic help in the practical phase of the study. We like to thank C.C. Delprat, MD and C.H. Statema, B.Sc., in particular, for their help in completing the data. We would also like to thank C.E.J.M. Kimmel, MD (FLEVO Ziekenhuis Almere, the Netherlands), R.C. Dirks, MD (Westfriesgasthuis Hoorn, the Netherlands) and R.E. Bernardus, MD (Tergooi Ziekenhuizen Blaricum, the Netherlands) for their help with the inclusion of patients.

FUNDING

This study was supported by a grant no. 171001003 from ZonMW, Organization for Health Research and Development, The Hague, The Netherlands. Molecular Biometrics Inc., USA, supplied the NIR spectroscopy technology. The funding sources did not influence the design and conduct of the study; the collection, management, analysis, and interpretation of the data; and the preparation, review or approval of the manuscript. The researchers were independent of the funders.

REFERENCES

1. Kjellberg AT, Carlsson P, Bergh C. Randomized single versus double embryo transfer: obstetric and paediatric outcome and a cost-effectiveness analysis. Hum Reprod 2006; 21: 210-216.

2. Källén B, Finnstrom O, Lindam A, Nilsson E, Nygren KG, Otterblad OP. Trends in delivery and neonatal outcome after in vitro fertilization in Sweden: data for 25 years. Hum Reprod 2010; 25: 1026–1034.

3. Thurin A, Hausken J, Hillensjö T, Jablanowska B, Pinborg A, Strandell A, Bergh C. Elective single-embryo transfer versus double-embryo transfer in in-vitro fertilization. N Engl J Med 2004; 351: 2392-2402.

4. Pandian Z, Bhattacharya S, Ozturk O, Serour G, Templeton A. Number of embryos for transfer following in-vitro fertilisation or intra-cytoplasmic sperm injection. Cochrane Database Syst Rev 2009; 15: CD003416.

5. Paternot G, Devroe J, Debrock S, D'Hooghe TM, Spiessens C. Intra- and inter-observer analysis in the morphological assessment of early-stage embryos. Reprod Biol Endocrin 2009; 29: 105-110.

6. Filho ES, Noble JA, Wells D. A review on automatic analysis of human embryo microscope images. Open Biomed Eng J 2010; 4: 170-177.

7. Kovalevsky G, Patrizio P. High rates of embryo wastage with use of assisted reproductive technology: a look at the trends between 1995 and 2001 in the United States. Fert Steril 2005; 84: 325-330.

8. Seli E, Robert C, Sirard MA. OMICS in assisted reproduction: possibilities and pitfalls. Mol Hum Reprod 2010; 16: 513-530.

9. Leese HJ. Metabolism of the preimplantation embryo: 40 years on. Reproduction 2012; 143: 417-427.

10. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. Proteomics 2006; 6: 4716-4723.

11. Scott RT, Seli E, Miller K, Sakkas D, Scott K, Burns DH. Noninvasive metabolomic profiling of human embryo culture media using Raman spectroscopy predicts embryonic reproductive potential: a prospective blinded pilot study. Fert Steril 2008; 90: 77-83.

12. Seli E, Sakkas D, Scott R, Kwok SC, Rosendhal SM, Burns DH. Noninvasive metabolomic profiling of embryo culture media using Raman and near-infrared spectroscopy correlates with reproductive potential of embryos in women undergoing in vitro fertilization. Fert Steril 2007; 88: 1350-1357.

13. Vergouw CG, Botros LL, Roos P, Lens JW, Schats R, Hompes PGA, Burns DH, Lambalk CB. Metabolomic profiling by near infrared spectroscopy as a tool to assess embryo viability: a novel, non-invasive method for embryo selection. Hum Reprod 2008; 23: 1499-1504.

14. Nagy ZP, Sakkas D, Behr B. Non-invasive assessment of embryo viability by metabolomic profiling of culture media ('metabolomics'). Reprod Biomed Online 2008; 17: 502-507.

15. Nagy ZP, Jones-Colon S, Roos P, Botros L, Greco E, Dasig J, Behr B. Metabolomic assessment of oocyte viability. Reprod Biomed Online 2009; 18: 219-225.

16. Seli E, Vergouw CG, Morita H, Botros LL, Roos P, Lambalk CB, Yamashita N, Kato O, Sakkas D. Noninvasive metabolomic profiling as an adjunct to morphology for noninvasive embryo assessment in women undergoing single embryo transfer. Fert Steril 2010; 94: 535-542.

17. Seli E, Bruce C, Botros LL, Henson M, Roos P, Judge K, Hardarson T, Ahlström A, Harrison P, Henman M, Go K, Acevedo N, Siques J, Tucker M, Sakkas D. Receiver operating characteristic (ROC) analysis of day 5 morphology grading and metabolomic Viability Score on predicting implantation outcome. J Assist Reprod Genet 2011; 28: 137-144.

18. Ahlström A, Wikland M, Rogberg L, Barnett JS, Tucker M, Hardarson T. Cross-validation and predictive value of near-infrared spectroscopy algorithms for day-5 blastocyst transfer. Reprod Biomed Online 2011; 22: 477-484.

19. Vergouw CG, Botros LL, Judge K, Henson M, Roos P, Kostelijk EH, Schats R, Twisk JWR, Hompes PGA, Sakkas D, Lambalk CB. Non-invasive viability assessment of day 4 frozen-thawed human embryos by near-infrared spectroscopy. Reprod Biomed Online 2011; 23: 769-767.

20. Hardarson T, Ahlström A, Rogberg L, Botros L, Hillensjö T, Westlander G, Sakkas D, Wikland M. Non-invasive metabolomic profiling of day 2 and 5 embryo culture medium: a prospective randomized trial. Hum Reprod 2012; 27: 89-96.

21. Sfontouris IA, Lainas GT, Sakkas D, Iliadis GS, Anagnostara K, Zorzovilis IZ, Petsas GK, Lainas TG. Assessment of embryo selection using non-invasive metabolomic analysis as an adjunct to morphology indicates improvement in implantation and fetal cardiac activity rates. Hum Reprod 2011; 26(Suppl 1): i86.

22. Mastenbroek S, Twisk M, van der Veen F, Repping S. Preimplantation genetic screening: a systematic review and meta-analysis of RCTs. Hum Reprod Update 2011; 17: 454-466.