Embryos cultured in a time-lapse system result in superior treatment outcomes; a strict matched pair analysis.
Title

Embryos cultured in a time-lapse system result in superior treatment outcomes; a
strict matched pair analysis.

Authors

Corresponding author:

amy.barrie@lwh.nhs.uk

A Barrie
The Hewitt Fertility Centre, Liverpool Women’s NHS Foundation Trust, Liverpool, UK

R Homburg
The Hewitt Fertility Centre, Liverpool Women’s NHS Foundation Trust, Liverpool, UK

G McDowell
Manchester Metropolitan University, Manchester, UK

J Brown
Edge Hill University, Ormskirk, UK

C Kingsland
The Hewitt Fertility Centre, Liverpool Women’s NHS Foundation Trust, Liverpool, UK

S Troup
The Hewitt Fertility Centre, Liverpool Women’s NHS Foundation Trust, Liverpool, UK

Word count excluding references: 2629
Abstract

A retrospective, strict matched-pair analysis on 728 treatment cycles between January 2011 and September 2014 was performed. 364 treatment cycles, where all embryos were cultured and examined in EmbryoScope®, were matched to treatment cycles where all embryos were cultured in a standard incubator with conventional morphological examination. Matching was performed for patient age, number of oocytes collected, treatment type and date of oocyte collection (± six months). The clinical (CPR), implantation (IR), live birth (LBR) and miscarriage rates (MR) were calculated and considered significant at p<0.05 (Chi-square test). CPR, IR and LBR were found to be statistically significantly higher in the time-lapse system (TLS) group compared to the standard incubation group (CPR: 44.8% versus 36.5%, p<0.03. IR: 39.3% versus 32.2%, p<0.03. LBR: 43.1% versus 33.8%, p<0.01). Although there was a 5.5% decrease in the MR for the TLS group when compared to the standard incubation group, this result was not statistically significant (18.9% versus 24.4%, p>0.1). There is a paucity of well-designed studies to confirm that embryos cultured and examined in TLS can result in superior treatment outcomes, and this strict-matched pair analysis with a large cohort of treatment cycles indicates the advantage of using TLS.

Key words: IVF, time-lapse, EmbryoScope, matched-pair
Introduction

The first application of time-lapse systems (TLS) in embryology was recorded in 1968 where chick embryos exposed to teratogenic doses of hypoxia were analysed (Grabowski and Schroeder, 1968). Following this studies relating to preimplantation embryonic development were published (Colly-d’Hooghe, Valleron, Malaise, 1977; Milligan, Harris, Dennis, 1978; Lueck and Aladjem, 1980; Massip and Mulnard, 1980; Milligan, Harris, Dennis, 1980; Schatten and Schatten, 1980; Alexandre and Mulnard, 1988). The first clinically relevant application of TLS for use in the embryology laboratory was reported in 1994 where the events of fertilisation were studied using a custom-made imaging system comprising a microscope, enabled with Nomarski DIC optics, a video recorder and a switching box (Inoué and Inoué, 1994). Regarding preimplantation embryonic development, the first report of internalisation of fragments was published (Hardarson et al., 2002) followed by a report of mouse embryo collapse analysed using time-lapse photography (Niimura, 2003). Focus turned to the use of TLS in a clinical setting, quite abruptly, in 2008 with a number of publications exclusively studying preimplantation embryonic development using TLS and how the information these systems provided could be used to determine embryo viability (Arav et al., 2008; Lemmen, Agerholm, Ziebe, 2008; Mio and Maeda, 2008). The first commercially available TLS began installations in Europe in 2011. TLS for clinical application have now been readily adopted worldwide with instruments installed in numerous countries. The body of evidence suggests that TLS can increase the chances of a pregnancy for many undergoing assisted reproduction. However, contradictory evidence exists. The use of TLS in clinical laboratories allows for a detailed analysis of embryos contained within it giving over 700 images per embryo compared to the conventional snap-shot observations translated into a
written series of numbers and letters open to interpretation by other members of the scientific team. The wealth of information that TLS provides inevitably creates the need to modify how embryos are selected for use and as such there are many reports linking time-lapse parameters (termed morphokinetics) to an embryos ability to create a pregnancy.

A recent Cochrane review retrieved 33 articles relating to the use of TLS with only ten studies being potentially eligible for inclusion (Armstrong et al., 2014). After further evaluation, three studies were included as true RCTs. These trials totalled 994 couples with the majority contributed by one study (856) (Rubio et al., 2012). Following analysis it was concluded that for all types of TLS, with or without cell-tracking, embryo-scoring algorithms, versus standard embryo incubation there was no conclusive evidence of a difference in clinical pregnancy, miscarriage, live birth, and stillbirth rates per couple randomised. The aim of the following investigation was to examine whether TLS can be considered superior to standard incubation systems when considering CPR, IR, LBR and MR by performing a strict matched-pair analysis with a large cohort of patients.

Methods

Study Design

A retrospective, observational, matched pair data analysis was designed and approved by the NHS Research Ethics Committee in the North West (ref: 14/NW/1043). Data for this research were obtained from 728 treatment cycles between January 2011 and September 2014. This data comprised 364 patients having embryos cultured in a standard incubator (Sanyo Multigas MCO-18M, 37°C, 6% carbon dioxide (CO₂)) (group 1) and 364 having their embryos cultured in a time-lapse enabled incubator, the EmbryoScope® (Vitrolife,
Gothenburg, 37°C, 6% CO₂, 5% oxygen (O₂) (group 2). Although in group 1 218 were cultured in 20% O₂ and 150 in 5% O₂, a thorough statistical examination of the primary outcome of this investigation showed no significant differences between these groups which were then pooled. All treatments included in this analysis were from known implantation embryos i.e. a single embryo transfer or a double embryo transfer where transfer of two embryos resulted in either a negative pregnancy test or two fetal heartbeats. Clinical pregnancy (CPR), implantation (IR), live birth (LBR) and miscarriage (MR) rates were calculated.

**Patient criteria**

All patients with embryos cultured in the EmbryoScope® with known outcome were matched to patients having embryos cultured in a standard incubator for patient age (exact), number of oocytes collected (exact), treatment type and date of treatment (± six months). Patients were not included twice. Patients were allocated to either standard or TLS culture randomly, based on availability. In February 2014 the laboratory became 100% time-lapse enabled meaning all patients had all embryos cultured in a TLS.

**Ovarian stimulation**

Pituitary down-regulation was achieved with either a gonadotrophin releasing hormone agonist (buserelin, Suprecur®, Sanofi-Aventis, UK) or antagonist (cetrorelix acetate, Cetrotide®, Merck Serono, Germany). Ovarian stimulation was performed using urine derived or recombinant follicle stimulating hormone (Progynova (Bayer, Germany), Fostimon, Merional (IBSA, Switzerland), Menopur® (Ferring Fertility, Switzerland), Gonal-f® (Merck Serono). Doses were adjusted based on patient demographic and response.


**Oocyte retrieval and embryology**

Ultrasound guided oocyte collection was performed transvaginally under sedation (Diprivan, Fresenius Kabi, USA). Collected oocyte-cumulus complexes were cultured in 4-well dishes (Nunc™, Thermo Scientific, USA) containing 0.65ml G-IVF™ (Vitrolife) covered with 0.35ml OVOIL™ (Vitrolife) in a standard incubator. Sperm preparation was performed using a standard gradient separation at 0.3 relative centrifugal force for ten minutes (ISolate®, Irvine Scientific, USA) followed by two washes at 0.6rcf for ten minutes using G-IVF™. Those oocytes destined for ICSI were prepared using enzymatic (HYASE-10XTM, Vitrolife) and mechanical digestion. ICSI was performed approximately four hours following collection after which time all injected oocytes were placed in individual culture drops of G-1™ (Vitrolife) and cultured in either the EmbryoScope® or a standard incubator. Those oocytes destined for standard insemination had this performed approximately four hours after collection and replaced into a standard incubator until fertilisation check the following day. Oocytes were then checked for fertilisation approximately 16-18 hours post-insemination (hpi) and all fertilised oocytes along with all unfertilised metaphase II oocytes were placed in individual culture drops of G-1™ and cultured in either the EmbryoScope® or a standard incubator. Where culture to day five was undertaken, a media change was performed on day three. For those embryos cultured in the EmbryoScope®, 20µl from each well was aspirated and replaced with 20µl of G-2™ (Vitrolife). For those embryos cultured in standard incubation, all embryos were moved to a new culture dish comprising individual 20µl drops of G-2™. Embryo selection for those cultured in the EmbryoScope® was performed using the national grading scheme (ACE/BFS guidelines, Cutting, Morroll, Roberts, Pickering, Rutherford, 2008) along with an internally derived embryo-scoring, time-lapse algorithm. Embryo
selection for those cultured in the standard incubator was performed using the national grading scheme only. Embryo transfer was performed using the highest-grade embryo(s) either three or five days post collection depending on the number of good quality embryos the patient had on day three as well as how many were to be transferred. Selected embryos were cultured in EmbryoGlue® (Vitrolife) prior to embryo transfer.

**Statistical analyses**

CPR was calculated as the number of patients having a fetal heart beat (fhb) at 6-7 weeks gestation confirmed by ultrasound scan (regardless of number of fbh) out of the number of embryo transfers performed. IR was calculated as the total number of fhb (i.e. inclusive of higher order pregnancies) out of the number of embryos transferred. LBR was calculated as the number of all live births out of the number of embryo transfers. Finally, MR was calculated as the number of positive human chorionic gonadotrophin (hCG) tests that did not result in a fhb at ultrasound scan at 6-7 weeks gestation. Results were analysed using the Chi-square test (GraphPad Software Inc).

**Results**

A total of 728 treatment cycles were analysed and the CPR, IR, LBR and MR calculated (see table 1 for data breakdown). CPR, IR and LBR were found to be statistically significantly different between the two groups (table 2). The CPR for group 1 (control, standard incubation) when compared to group 2 (TLS) was 36.5% versus 44.8%, respectively, p < 0.03. The IR for group 1 when compared to group 2 was 32.2% versus 39.3%, respectively, p < 0.03. The LBR for group 1 when compared to group 2 was 33.8% versus 43.1%, respectively, p < 0.01. However, although there was a 5.5% increase
in the MR for group 1 when compared to group 2, this result was not statistically significant (24.4% versus 18.9%, respectively, p > 0.1).

[table 1 near here]

[table 2 near here]

Discussion

The results of this matched pair analysis reveal that embryos cultured and examined in the EmbryoScope® incubator result in superior treatment outcomes. These results are in concordance with others (Rubio et al., 2014; Yang et al., 2014; Adamson et al., 2015; Basile et al., 2015) but have been contradicted elsewhere (Nakahara et al., 2010; Cruz et al., 2011; Kirkegaard et al., 2012; Kahraman et al., 2013; Kovacs et al., 2013; Armstrong et al., 2014; Park et al., 2015). A recent Cochrane review (Armstrong et al., 2014) suggested that there was insufficient evidence to conclude that TLS with or without cell tracking technology would be beneficial to patients undergoing ART. Included in this were three eligible RCTs the first of which contributed most of the data for the review. This study was a multi-centre RCT of patients undergoing ICSI, using donated or autologous oocytes. 856 couples were randomised in this analysis; 444 to TLS and 412 to standard incubation. The CPR and MR were calculated as end-points. Although this analysis revealed a significant increase in treatment outcomes, considerable reasons for bias were identified. Firstly, patients could request the intervention (TLS) therefore allocation was, in fact, non-random. Secondly, the study was classed as ‘double-blinded’ due to the gynaecologist and statistician being unaware of the arm to which the patients had been randomised. However, the patients and embryologists were given this information. Although unlikely to create a significant bias, this detail could invalidate the results. Finally, the heterogeneity of the sample was considerable including the use of
donated, and thus both fresh and frozen oocytes (Rubio et al., 2014). The remaining two RCTs included in the review were conducted on a small number of couples, one being interim results only, leaving a combined total of 61 to 65 in each arm (Kahraman et al., 2013; Kovacs et al., 2013). The reviewers reported a high risk of attrition bias in one of these studies due to the principal investigator undertaking the randomisation and also because there was no blinding. Overall, the reviewers stated that there was no conclusive evidence of a difference between standard incubation and TLS when considering CPR, MR, LBR and stillbirth rates. Further analyses, not included in this Cochrane review, have also shown no significant differences in treatment outcomes between embryos cultured in standard incubation versus TLS (Nakahara et al., 2010; Cruz et al., 2011; Kirkegaard et al., 2012; Park et al., 2015).

Differences in results found thus far in the matter of TLS could be attributed to a number of factors. Firstly, a benefit of TLS that one laboratory might enjoy may not be so with another due to the conditions of the laboratory in the first instance. In brief, a well-designed, stable culture environment (TLS) introduced into what was a relatively unstable culture condition may elicit an immediate uplift in treatment outcomes. Whereas, to place this technology into an already optimal culture environment, may not reveal such results. There are many factors that vary between laboratories that could impact this; the type of culture media (single or sequential), culture dish type, volume of media used for culturing embryos, volume of oil overlay, the type of incubator and the embryo grading and embryo transfer policies. It is therefore reasonable to conclude that some laboratories may benefit from TLS more than others.
Secondly, during the culture of embryos in the EmbryoScope® in this analysis an in-house derived embryo-scoring algorithm was used. This indicates that the analysis presented here does not distinguish between the two, commonly stated, major benefits of TLS; the undisturbed nature of the systems or the use of embryo-scoring algorithms. Whilst the authors acknowledge that this could create ambiguity, it can also be defended. This detail means that this analysis addresses TLS as a whole in the manner in which it should be utilised; using the information provided by the images. It also gives further explanation for the heterogeneity of success of TLS. Some laboratories utilising TLS have access to large amounts of data meaning in-house derivation and validation of predictive models can be performed; a method much preferred to utilising a published embryo-scoring algorithm developed externally. In these laboratories, where internally derived models are used, although not proven, a greater benefit to using TLS would be expected. Naturally, in those laboratories that do not have access to a dataset allowing in-house derivation of predictive models, externally developed versions may be adopted, a decision which has been cautioned (Kirkegaard et al., 2013a; Yalçınkaya et al., 2014). Evidence of the benefits to using embryo-scoring algorithms can be seen clearly in the literature where the earliest publications regarding TLS aimed to assess the safety of the systems (Nakahara et al., 2010; Cruz et al., 2011; Freour et al., 2012; Kirkegaard et al., 2012). Many of these studies randomised oocytes or embryos between two culture systems (standard and TLS) and found no differences in treatment outcomes of embryo quality parameters. The use of an embryo-scoring algorithm in these studies is not mentioned, thus these analyses assessed the effectiveness of the incubator itself, not the information it provided. Once satisfaction with the safety of the system had been reached attention was then turned to how the information from the TLS could be utilised. Further reports were then published that revealed an uplift in outcome parameters (Rubio et al., 2014;
VerMilyea et al., 2014; Yang et al., 2014; Basile et al., 2015; Milewski et al., 2015; Siristatidis et al., 2015) with one obvious difference; these analyses included the use of an embryo-scoring algorithm. A particular study, that evidences the above notion, randomised 843 couples undergoing ICSI in a double-blinded manner to either standard incubation or TLS. Those embryos cultured in a standard incubator were assessed for selection based on morphology alone whereas those in TLS were selected using an internally derived, multivariable model. A significantly higher ongoing pregnancy rate was found in TLS compared to standard incubation (51.4% versus 41.7% per cycle and 54.5% versus 45.3% per embryo transfer, respectively) as well as a significantly decreased early pregnancy loss in TLS (16.6% versus 25.8%). In addition, the implantation rate was significantly increased in the TLS group (44.9% versus 37.1%) (Rubio et al., 2014). A further investigation sought to select the most competent blastocysts for transfer by combining TLS and array comparative genomic hybridisation (aCGH) for patients undergoing preimplantation genetic screening (PGS) designed as a prospective study with sibling oocytes. 1163 metaphase II oocytes from 138 PGS patients were included and oocytes were randomised to two groups after ICSI; group A were cultured in TLS and group B in standard incubation. Array CGH using trophectoderm biopsy on both groups was carried out and one or two euploid blastocysts either within the morphokinetic ranges (group A) or morphological grades (group B) were transferred. The clinical pregnancy and implantation rates were found to be significantly higher in group A when compared to group B (CPR; 71.1% versus 45.9%, IR; 66.2% versus 42.4%, respectively) demonstrating that when embryo-scoring algorithms are used as an adjunct to select embryos for transfer, superior treatment outcomes can be achieved (Yang et al., 2014). These investigations address TLS as an incubator whilst also using the data it provides, synonymous with the current analyses, indicating that embryo-scoring
algorithms derived using TLS are able to select embryos more effectively than standard morphology assessments. It is not surprising that an increase in treatment outcomes is seen in these cases owing to the wealth of information that is available to the user of TLS to do basic, but powerful, embryo de-selection.

Literature regarding TLS now predominantly concerns development of embryo-scoring algorithms or reviews concluding that further evidence for its (TLS) superiority is required. The authors believe that predictive models can be very useful, in the first instance for de-selection of embryos undergoing abnormal cleavage events such as direct cleavage and reverse cleavage shown to have a significantly reduced chance of creating a pregnancy (Rubio et al., 2012; Liu et al., 2014) but also, if developed effectively, to select the best embryo from a cohort for a specific patient demographic. Patient characteristics including infertility diagnosis (Sundvall et al., 2015) and maternal age (Hampl and Stephan, 2013; Chawla et al., 2015) as well as treatment characteristics including treatment type and culture conditions (Lemmen et al., 2008; Wale and Gardner, 2010; Ciray et al., 2012; Cruz et al., 2013; Kirkegaard et al., 2013b) have been shown to affect an embryo’s morphokinetic profile and the resulting subtle differences may be used to identify which embryo has the highest implantation potential. Herein lies a further reason for possible variation in success of TLS between laboratories; patient and treatment characteristics.

**Conclusion**

This matched pair analysis indicates that treatment cycles where embryos are cultured and examined in TLS result in superior outcomes including clinical pregnancy, implantation, live birth and miscarriage rates. Although the notion is novel, the authors believe that the
real benefit of TLS lies in the development of patient specific embryo-scoring algorithms. Literature thus far indicates that there is likely to be no difference in treatment outcomes when an embryo-scoring algorithm is not used and future research should be geared towards developing effective embryo-scoring algorithms to aid in embryo selection.

Declaration of interest
The authors report no declaration of interest.

References


Grabowski CT, Schroeder RE. A time-lapse photographic study of chick embryos exposed to teratogenic doses of hypoxia. Exp Hematol. (1968)19(3): 347-62


<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single embryo transfers (SET, n)</td>
<td>300</td>
<td>283</td>
<td>0.11</td>
</tr>
<tr>
<td>Double embryo transfers (DET, n)</td>
<td>64</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>Cleavage stage transfers (n)</td>
<td>116</td>
<td>101</td>
<td>0.20</td>
</tr>
<tr>
<td>Blastocyst stage transfers (n)</td>
<td>248</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td>Previous attempts (mean ± st.dev.)</td>
<td>1.35 (±0.92)</td>
<td>1.45 (±0.95)</td>
<td></td>
</tr>
<tr>
<td>Embryos transferred (n)</td>
<td>428</td>
<td>445</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Table 1: baseline data indicating the number of SET and DET, proportion of cleavage and blastocyst transfers, number of previous attempts and the number of embryos transferred in group 1 and group 2. Data were analysed using the Chi-square test.
<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Standard)</td>
<td>(EmbryoScope®)</td>
<td></td>
</tr>
<tr>
<td>Clinical pregnancy rate (%)</td>
<td>133/364 (36.5)</td>
<td>163/364 (44.8)</td>
<td>&lt;0.03*</td>
</tr>
<tr>
<td>Implantation rate (%)</td>
<td>138/428 (32.2)</td>
<td>175/445 (39.3)</td>
<td>&lt;0.03*</td>
</tr>
<tr>
<td>Live birth rate (%)</td>
<td>123/364 (33.8)</td>
<td>157/364 (43.1)</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Miscarriage rate (%)</td>
<td>43/176 (24.4)</td>
<td>38/201 (18.9)</td>
<td>&gt;0.1</td>
</tr>
</tbody>
</table>

**Table 2**: data end point results. CPR, IR, LBR and MR for both standard and EmbryoScope® incubation. All results were considered to be statistically significantly different between the two groups where p<0.05 (Chi-square test).