Approaches to improve the diagnosis and management of infertility

P. Devroey1,4, B.C.J.M. Fauser2 and K. Diedrich3 on behalf of the Evian Annual Reproduction (EVAR) Workshop Group 2008†

1Center of Reproductive Medicine, Free University Brussels, Laarbeeklaan 101, 1090 Brussels, Belgium 2Department of Reproductive Medicine and Gynaecology, University Medical Center, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands 3Department of Obstetrics and Gynecology, University Clinic of Schleswig-Holstein, Campus Luebeck, 23538 Luebeck, Germany

4Correspondence address. Tel: +32-2-477-66-60; Fax: +32-2-477-66; E-mail: paul.devroey@uzbrussel.be

TABLE OF CONTENTS

- Introduction
- Methods
- Individualized pretreatment assessment
- Management strategy selection
- Individualized ovarian stimulation
- Embryo assessment
- Laboratory standards and procedures
- Single embryo transfer
- Evaluation of ART treatment outcomes
- The future of ART
- Conclusions

BACKGROUND: Recent advances in our understanding of the causes of infertility and of assisted reproductive technology (ART) have led to the development of complex diagnostic tools, prognostic models and treatment options. The Third Evian Annual Reproduction (EVAR) Workshop Meeting was held on 26–27 April 2008 to evaluate evidence supporting current approaches to the diagnosis and management of infertility and to identify areas for future research efforts.

METHODS: Specialist reproductive medicine clinicians and scientists delivered presentations based on published literature and ongoing research on patient work-up, ovarian stimulation and embryo quality assessment during ART. This report is based on the expert presentations and subsequent group discussions and was supplemented with publications from literature searches and the authors’ knowledge.

RESULTS: It was agreed that single embryo transfer (SET) should be used with increasing frequency in cycles of ART. Continued improvements in cryopreservation techniques, which improve pregnancy rates using supernumerary frozen embryos, are expected to augment the global uptake of SET. Adaptation and personalization of fertility therapy may help to optimize efficacy and safety outcomes for individual patients.

†The Third Evian Annual Reproduction (EVAR) Workshop Meeting (26–27 April 2008) was organized to discuss personalized approaches to the diagnosis and management of infertility and was supported by an unrestricted educational grant from Merck Serono S.A.—Geneva. The content of this manuscript is based on the presentations and discussions of the EVAR Workshop Meeting. The meeting was chaired by P.D. (Free University Brussels, Brussels, Belgium), K.D. (University Clinic of Schleswig-Holstein, Luebeck, Germany) and B.C.J.M.F. (University Medical Center, Utrecht, The Netherlands). The speakers included Carlo Alviggi (Università degli Studi di Napoli Federico II, Naples, Italy), Esther Baart (University Medical Center, Utrecht, The Netherlands), Christopher Barrat (Ninewells Hospital, Dundee, UK), Christina Bergh (Saligreniska University Hospital, Gothenburg, Sweden), Frank Broekmans (University Medical Center, Utrecht, The Netherlands), Johannes L.H. Evers (Maastricht University Medical Center, Maastricht, The Netherlands), Georg Griesinger (University Clinic of Schleswig-Holstein, Luebeck, Germany), Stephen Keay (University of Warwick, Coventry, UK), François Olivennes (ART unit Eylau La Muette, Paris, France), Gamal Serour (Al-Azhar University, Cairo, Egypt), Catherine Staessen (Vrije Universiteit, Brussels, Belgium), Arne Sunde (St Olav’s University Hospital in Trondheim, Norway), Christos Venetis (Aristotle University of Thessaloniki, Greece) and Hakan Yarali (Hacettepe University, Ankara, Turkey). Colin M. Howles (Merck Serono S.A.—Geneva) and Julian Jenkins (previously of Merck Serono S.A.—Geneva) were in attendance.
patients. Prognostic modelling and personalized management strategies based on individual patient characteristics may prove to represent real progress towards improved treatment. However, at present, there is limited good-quality evidence to support the use of these individualized approaches.

**CONCLUSIONS:** Greater quality control and standardization of clinical and laboratory evaluations are required to optimize ART practices and improve individual patient outcomes. Well-designed, good-quality studies are required to drive improvements to the diagnosis and management of ART processes.

**Key words:** infertility work-up / in vitro fertilization / ovarian response / predictive factors / single embryo transfer

---

**Introduction**

Infertility can be defined as the failure to achieve a pregnancy within 1 year of regular unprotected intercourse (Evers, 2002; Zegers-Hochschild et al., 2006). Despite the inherent difficulties of estimating the prevalence of infertility (Greenhall and Vessey, 1990), it is generally accepted that one in four women are affected at sometime (Gunnell and Ewings, 1994). Moreover, ~20% of couples consult their general physician because of difficulty conceiving, and half of those couples (10%) require specialist care (Hull et al., 1985; Beurskens et al., 1995).

Assisted reproductive technology (ART) treatment for infertile couples now results in reasonably high pregnancy rates but may be associated with risks of ovarian hyperstimulation syndrome (OHSS) and multiple pregnancy (Van Voorhis, 2006). The variability in patient characteristics and response to ART dictate the need for proven, personalized diagnostic and therapeutic approaches to optimize efficacy and safety outcomes (Fauser et al., 2008).

Advances in our understanding of the causes of infertility and ART have facilitated the development of increasingly complex diagnostic tools, prognostic models and treatment options. It is hoped that the identification of reliable baseline demographic, disease or genetic characteristics that are predictive of treatment outcome will enable selection of the most appropriate management strategy for each couple. Furthermore, new laboratory techniques may facilitate the transfer of single embryos while maintaining existing pregnancy rates (Van Voorhis, 2006).

Personalized management strategies, based on individual patient characteristics, have been proposed and the further development of these strategies may represent real progress towards individually tailored fertility treatment. The Third EVAR Annual Reproduction (EVAR) Workshop Meeting was held on 26–27 April 2008 to evaluate the existing evidence to support current approaches to the diagnosis and management of infertility. Here, we report the discussions and expert opinions of the EVAR Workshop Group based on presentations summarizing current literature and identify key areas for future research efforts.

---

**Methods**

Prior to the Third EVAR Workshop Meeting in April 2008, specialist reproductive medicine clinicians and scientists prepared presentations based on published literature and ongoing research relating to patient work-up, ovarian stimulation and embryo quality assessment during ART. Following the presentations, group discussion to reach joint conclusions on the topics covered was facilitated by the chairmen: P.D., B.C.J.M.F. and K.D.

The content of this report is based on the expert presentations and subsequent group discussions that took place during the workshop meeting. Given the broad scope of this report, a systematic literature review was not feasible. Instead, the discussions relating to each topic were complemented with electronic literature searches via PubMed for articles of any type that were published in the English language and limited by date of publication. Combinations of the following keywords were used to identify relevant articles: ‘assisted reproduction’, ‘ART’, ‘blastocyst’, ‘cryopreservation’, ‘embryo’, ‘FSH’, ‘ICSI’, ‘individualized’, ‘infertility’, ‘IU’, ‘IVF’, ‘LH’, ‘multiple pregnancy’, ‘outcome’, ‘ovarian’, ‘PGS’, ‘post-coital test’, ‘prognostic’, ‘reserve’, ‘resistance’, ‘response’, ‘sperm analysis’, ‘SET’, ‘stimulation’, ‘tubal occlusion’ and ‘work-up’. The literature search was also supplemented with key publications that were known to the authors.

**Individualized pretreatment assessment**

The accurate detection of underlying reproductive abnormalities helps to guide individual management decisions and maximize ART treatment outcomes. Clinical evaluation of the infertile couple may be grouped into five categories: semen analysis, the post-coital test (PCT), assessment of ovulation, uterine and tubal evaluation, and laparoscopy (Balasch, 2000). Of these, semen analysis, mid-luteal phase serum progesterone level and tubal patency evaluation comprise the initial basic patient work-up (Crosignani and Rubin, 2000). However, the use of several fundamental elements of infertility testing is still contentious, and evidence suggests that the current World Health Organization (WHO) recommendations for the standard investigation of the infertile couple are poorly followed in Europe (Rowe et al., 1993; Balasch, 2000).

**Semen analysis**

Humans have a low proportion of ‘normal’ sperm compared with many other species. Although relatively few studies of semen analysis have been performed in men with proven fertility, there is a high degree of overlap in semen characteristics between fertile and infertile men (Guzick et al., 2001). High-quality semen analysis has diagnostic value for gross male infertility conditions (such as azoospermia or globozoospermia), but the predictive value of an individual semen analysis is less robust when moderate numbers of motile sperm are present (Comhaire, 2000).

Semen analysis comprises sperm concentration, motility and morphology. No isolated semen analysis measures have been shown to be diagnostic of infertility in large studies (Guzick et al., 2001). In an effort to increase the value of semen analyses, results have been
incorporated into complex prediction models (Snick et al., 1997; Hunault et al., 2004). However, the output of these models has large confidence intervals and results must be interpreted cautiously (Snick et al., 1997; Hunault et al., 2004).

There are various methods for semen analysis, but those recommended by the WHO and the European Society for Human Reproduction and Embryology (ESHRE) provide the most comprehensive and robust methods (World Health Organization, 1999; Kvist and Bjorndahl, 2002). Although a new manual is due to be published in 2009, the WHO currently defines ‘normal’ as a sperm concentration of $>20 \times 10^6$/ml with $>50\%$ of progressively motile sperm and the presence of sperm with standard (no parameters provided) morphology (World Health Organization, 1999).

Evidence suggests that the WHO recommendations for performance of semen analysis and reporting of results are adhered to poorly in routine laboratory practice (Keel et al., 2002; Riddell et al., 2005). Despite the availability of established systems to improve staff training in semen assessments, such as ESHRE courses (Bjorndahl et al., 2002), the majority of laboratories still do not have accurate methods or appropriate training systems. Thus, semen analysis results are often variable. The demonstrated absence of standardization and strict quality control for semen analysis undermines the diagnostic and prognostic value of the test.

Despite the limitations described, semen analysis is routinely used to evaluate the fertilization potential of the male partner in infertile couples. Semen analysis outcomes also guide management decisions and often influence the choice of expectant management, intruterine insemination (IUI), in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI).

Greater standardization of semen analysis and accurate laboratory evaluation is clearly needed to improve the prognostic value of semen analysis (Ombelet et al., 2003). Furthermore, high-quality studies are required to identify threshold levels that are predictive of treatment outcome to assist decision-making for ART treatment. Sperm function tests may offer greater predictive power than traditional semen analysis but require strict validation prior to use in routine clinical practice (Lefievre et al., 2007).

**Post-coital test**

The PCT provides an assessment of the quantity and quality of cervical mucus, sperm–mucus interactions and the presence of antisperm antibodies (Balasch, 2000; Glazener et al., 2000). The test involves microscopic examination of extracted endocervical mucus, which should be conducted in the pre-ovulatory phase and 8–12 h after intercourse (Glazener et al., 2000). A positive PCT result may be defined as the presence of at least one forward-progressing spermatozoon in more than half of at least five high-power ($\times 400$ magnification) microscope fields (Glazener et al., 2000).

The PCT may provide an effective predictor of conception for couples with an infertility history of $<3$ years and no identified female causes of infertility (Glazener et al., 2000). However, it has been shown that the outcome of the PCT can be predicted for only half of all infertile couples (in which the female partner has a regular menstrual cycle) using the basic elements of an infertility history and semen analysis results (van der Steeg et al., 2004). Furthermore, results of the PCT are subject to considerable intra- and inter-observer variability (Male Infertility Best Practice Policy Committee of the American Urological Association and Practice Committee of the American Society for Reproductive Medicine, 2006). Given the widespread use of ART for couples with a negative PCT, routine use of this test in clinical practice is not recommended (National Institute for Clinical Excellence, 2004; Male Infertility Best Practice Policy Committee of the American Urological Association and Practice Committee of the American Society for Reproductive Medicine, 2006).

**Uterine and tubal evaluation**

Evaluation of the morphology of the uterus and fallopian tubes is essential prior to initiating ART (Crosignani and Rubin, 2000). Uterine and tubal structures may be visualized using a variety of techniques, including hysteroscopy, hysterosalpingography (HSG), transvaginal ultrasonography (TVS), saline infusion sonohysterography (SIS), magnetic resonance imaging (MRI) or laparoscopy.

Tubal occlusion is estimated to be the cause of infertility for 14% of couples who require specialist treatment (Hull et al., 1985). The optimum screening strategy to assess tubal damage is a complex subject with currently no consensus of opinion (National Institute for Clinical Excellence, 2004; den Hartog et al., 2008). A recent observational study comparing screening strategies for tubal factor infertility with laparoscopy has provided valuable data (den Hartog et al., 2008). Chlamydia trachomatis immunoglobulin G serology was shown to discern accurately patients at high versus low risk of tubal pathology and, thus, was suggested to provide a useful initial screening test (den Hartog et al., 2008). HSG was worthwhile only for low-risk patients to confirm the absence of tubal pathology and confer potentially beneficial effects of tubal flushing with oil-soluble contrast medium (den Hartog et al., 2008). Laparoscopy with dye testing was considered useful for patients with positive Chlamydia serology or evidence of bilateral tubal occlusion on HSG (den Hartog et al., 2008).

The presence of hydrosalpinx has a negative effect on clinical pregnancy and live birth rates following IVF (Strandell, 2007). The mechanism by which a hydrosalpinx impairs IVF success is incompletely understood, but its removal prior to IVF is known to significantly improve treatment outcomes (Strandell, 2007). A Cochrane meta-analysis of data from three clinical studies demonstrated superior pregnancy [odds ratio (OR) 1.75, 95% confidence interval (CI) 1.07–2.86] and live birth rates (OR 2.13, 95% CI 1.24–3.65) following salpingectomy compared with no surgical intervention (Johnson et al., 2004). Evidence from one of these studies suggested that the benefit of salpingectomy is greatest for patients with fluid-filled hydrosalpinges that are visible on ultrasound examination (Strandell et al., 1999). Thus, patients with a hydrosalpinx that is visible on ultrasound should be encouraged to undergo prophylactic salpingectomy prior to IVF.

The optimal management of endometriotic ovarian cysts in infertile patients is less well defined. Recent evidence of reduced responsiveness to gonadotrophins following laparoscopic ovarian cystectomy has challenged the traditional surgical approach to treatment (Somigliana et al., 2006). Indeed, it has been suggested that surgery should be undertaken only for the treatment of large endometriomas or pain that is refractory to medical treatment, or to exclude malignancy (Garcia-Velasco and Somigliana, 2009). Unexpected hysteroscopic abnormalities have been reported in up to 40% of patients during ART work-up (Shamma et al., 1992; Doldi...
et al., 2005), but there is no compelling evidence that either routine use of hysteroscopy before IVF or correction of identified pathology leads to better treatment outcomes. Compared with hysteroscopy, HSG has high sensitivity (81–98%) but low specificity (23–35%), and high false-negative (10–90%) and false-positive (22–44%) rates.

Although hysteroscopy is considered the gold standard for identification of intrauterine pathology (Bozdag et al., 2008), recent advances have enabled ultrasonographic techniques to increasingly substitute for invasive screening procedures (Ekerhovd et al., 2004). Late follicular phase TVS has proved to be a useful tool for the detection of intrauterine abnormalities such as polyps, synchieae, fibroids, and Mullerian anomalies (Van Voorhis, 2008). SIS offers enhanced visualization of the endometrium and better detection of intrauterine pathology than does standard TVS, and may be as effective as hysteroscopy in detecting intracavitary abnormalities (Ragni et al., 2005; Valenzano et al., 2006). MRI may be used for patients with suspected complex Mullerian anomalies (Deutch and Abuhamad, 2008).

Uterine fibroids occur in 20–50% of women aged over 30 years and are the most common benign tumour of the female genital tract (Eldar-Geva et al., 1998; Okolo, 2008). These tumours are heterogeneous in composition, size, location and number (Pritts, 2001), thus complicating the identification of women who would benefit from myomectomy prior to ART treatment.

Retrospective data suggest that the presence of fibroids <4 cm in diameter does not affect the outcome of ART treatment cycles (Vimercati et al., 2007). Additional retrospective analyses suggest that only fibroids that encroach on the uterine cavity negatively affect implantation rates and pregnancy outcomes in ART (Farhi et al., 1995; Eldar-Geva et al., 1998). A meta-analysis suggested that compared with infertile women without fibroids, women with submucosal fibroids have significantly lower rates of clinical pregnancy [relative risk (RR) 0.36, 95% CI 0.18–0.74], implantation (RR 0.28, 95% CI 0.12–0.65) and live birth (RR 0.32, 95% CI 0.12–0.85) (Pritts, 2001). Removal of submucosal fibroids improves clinical pregnancy rates (RR 2.03, 95% CI 1.08–3.83), but the limited available data suggest no improvement in treatment outcomes after removal of intramural fibroids (Pritts, 2001).

Endometrial polyps have been identified by hysteroscopy in 16–27% of women with otherwise unexplained infertility (Kim et al., 2003; de Sa Rosa e de Silva et al., 2005). The benefit of hysteroscopic polypectomy on pregnancy rate has been demonstrated in a prospective, randomized study of women with ultrasonically diagnosed endometrial polyps who were undergoing IUI (Perez-Medina et al., 2005), in which patients who underwent polypectomy had a significantly higher cumulative pregnancy rate than those who underwent hysteroscopy plus polyp biopsy (63.4% versus 28.2%, \( P < 0.001 \)) (Perez-Medina et al., 2005).

An association between polypectomy and improved spontaneous pregnancy rates was also shown in five non-randomized studies (Varasteh et al., 1999; Spiewankiewicz et al., 2003; Shokeir et al., 2004; Stamatellos et al., 2008; Yanaihara et al., 2008). Retrospective data suggest that hysteroscopic polypectomy improves pregnancy rates in previously infertile women, regardless of the number or size of polyps present (Stamatellos et al., 2008), and that resection of polyps located at the utero-tubal junction may improve pregnancy rates in infertile patients (Yanaihara et al., 2008). Although the effect of endometrial polyps on IVF is unclear (Lass et al., 1999; Isikoglu et al., 2006), data suggest that women with otherwise unexplained infertility may still benefit from polypectomy (Stamatellos et al., 2008).

In summary, prophylactic salpingectomy improves ART outcomes for patients with a fluid-filled hydrosalpinx. However, there are currently insufficient high-quality data on the optimum screening modality and management of other uterine and tubal abnormalities prior to ART on which to base personalized patient care. Good-quality, prospective studies are warranted to evaluate the relative merits of uterine and tubal screening tests and management approaches prior to ART treatment.

**Evaluation of ovarian reserve**

Ovarian stimulation is used in ART to stimulate multifollicular development and enable multiple oocyte retrieval (Fauser et al., 2008). The ovarian response to gonadotrophin stimulation depends primarily on the woman’s ovarian reserve (Broekmans et al., 2006) and is a major determinant of the success of IVF (van der Gaast et al., 2006; Shanbhag et al., 2007).

Ovarian reserve represents the remaining population of primordial and resting follicles (Gougeon, 1996) and is generally defined as the quantity and quality of the follicles present in the ovary (Broekmans et al., 2006). For operational purposes, the ovarian reserve can be defined as the number of antral follicles present in the ovaries at a given time that can be stimulated into dominant follicle growth by exogenous follicle-stimulating hormone (FSH). Women with a so-called ‘normal’ ovarian reserve will develop an average of 8–10 dominant follicles in response to conventional ovarian stimulation, with a corresponding number of oocytes (Broekmans et al., 2006).

Although chronological age is the major determinant of ovarian reserve, there is considerable individual variability in the rate of ovarian ageing (Fig. 1) (te Velde and Pearson, 2002). Therefore, accurate tests of ovarian reserve may allow individualized predictions of oocyte yield and ART treatment outcome in terms of ongoing pregnancy (Broekmans et al., 2006).

![Figure 1](https://academic.oup.com/humupd/article-abstract/15/4/391/734669/1544)  
*Figure 1* Variation in reproductive ageing.  
Adapted with permission from te Velde and Pearson (2002). Curve 1 shows the Gaussian distribution of variation of age at menopause, Curve 2 the variation of age of transition from cycle regularity to irregularity, Curve 3 the variation in age of becoming sterile and Curve 4 the variation in age of becoming infertile.
Serum FSH level and antral follicle count (AFC) assessed by TVS are often used as tests of ovarian reserve (Broekmans et al., 2006). The AFC correlates with the number of oocytes retrieved and ART outcomes (Frattarelli et al., 2003) and is currently considered to be the best available single predictor of ovarian response to stimulation (Bancsi et al., 2002; Hendriks et al., 2007). Serum anti-Müllerian hormone (AMH) levels may accurately identify patients at risk of an extreme ovarian response (Nelson et al., 2007), but are infrequently measured in routine practice (van Rooij et al., 2002; van Rooij et al., 2005; Broer et al., 2009).

The evaluation of ovarian reserve may enable the identification of patients (with regular menstrual cycles) who will have a better or worse response to gonadotrophin stimulation than would be expected for their chronological age (Broekmans et al., 2006). In theory, this would help clinicians to personalize patient management by selecting an appropriate treatment protocol, stressing the need for early initiation of treatment or counselling against initiation of treatment (Broekmans et al., 2006).

Studies on the impact of ovarian reserve tests to select the appropriate starting dose of FSH have yielded contradictory results (Popovic-Todorovic et al., 2003a, b; Klinkert et al., 2005; Olivennes et al., 2009).

Although the potential use of AMH needs to be evaluated further, all widely available tests of ovarian reserve are poor predictors of clinical pregnancy and live births after IVF (Hendriks et al., 2005; Mol et al., 2006). As such, a patient’s true ovarian reserve can be determined only after a cycle of ovarian stimulation.

Management strategy selection

Expectant or active therapy

Infertile couples can be divided into two groups: those who are unable to conceive without therapy and those who have reduced fertility but are likely to conceive spontaneously with time (Crosignani and Rubin, 2000). Expectant management is the most appropriate approach for infertile couples with a good prognosis for spontaneous pregnancy, whereas ART provides a valuable treatment option for selected couples with a low probability of spontaneous pregnancy (Broekmans et al., 2006). A reliable estimate of the likelihood of spontaneous pregnancy is essential to enable clinicians to decide whether expectant or active management is more appropriate for a given couple (Hunault et al., 2004).

Prognostic models derive the likelihood of spontaneous pregnancy in an individual couple based on large-sample data (Comhaire, 1987; Eimers et al., 1994; Wichmann et al., 1994; Collins et al., 1995; Snick et al., 1997). Statistical analysis of a model developed by Snick et al. (1997) demonstrated that it would be predictive of a live birth in 76–79% of infertile couples in a primary care setting. The most important prognostic factors in this model were an abnormal PCT, a tubal defect, an ovulatory defect and infertility of longer than 2 years (Snick et al., 1997). Although female age is considered to be one of the most important factors to affect ART outcome (Templeton et al., 1996; van Kooij et al., 1996), female age (≤30 years) was predictive of live birth in this model only when the PCT result was excluded (Snick et al., 1997).

The use of such prognostic models in clinical practice requires standardized screening assessments for all patients. Furthermore, there are no strict criteria on which to base management decisions. Hence, the likelihood of spontaneous pregnancy for each individual couple must be weighed against the potential benefits or risks of interventional treatment.

Active ART treatment options

IVF is an effective treatment option for female infertility, whereas ICSI was developed for male infertility. Here, we discuss evidence to support the use of these treatment options in cases of male infertility, bilateral tubal occlusion and unexplained infertility.

Male infertility

Couples affected by severe male infertility related to conditions such as obstructive or non-obstructive azoospermia require treatment with ICSI to achieve pregnancy (National Institute for Clinical Excellence, 2004). However, mild-to-moderate male infertility is a poorly defined concept and treatment strategies are highly variable.

Limited evidence suggests that IUI after clomiphene citrate stimulation (CC-IUI) may be an effective first-line therapy for male infertility when the total inseminating motile sperm count after preparation is >1 x 10⁶/ml (Ombrelet et al., 2003). Furthermore, CC-IUI still represents an effective therapeutic option when the inseminating motile sperm count is <1 x 10⁶/ml if the sperm morphology score is at least 4% (Van Waart et al., 2001; Ombrelet et al., 2003). Despite these findings, the authors of a recent Cochrane meta-analysis concluded that there is insufficient evidence to support the use of IUI rather than timed intercourse (with or without ovarian stimulation) for male infertility, and called for additional large, high-quality, randomized controlled trials to investigate this issue (Bensdorp et al., 2007).

There are also limited data to support the use of conventional IVF or ICSI for mild-to-moderate male infertility. Clinical pregnancy rates were similar after up to six cycles of IUI (with or without ovarian stimulation) or IVF for male infertility in a prospective, randomized study (Goverde et al., 2000). A meta-analysis of data from nine randomized, controlled, sibling oocyte design trials of IVF versus ICSI for couples with ‘borderline semen characteristics’ demonstrated a pooled relative fertilization rate of 2.2 (95% CI 1.6–3.0) per oocyte in favour of ICSI (Tournaye et al., 2002). However, fertilization rates were not significantly different in a sibling oocyte design study of ICSI or high insemination concentration (0.8 sperm x 10⁶/ml) IVF (67.6 versus 59.6; P = 0.066) (Tournaye et al., 2002). There is a need for high-quality trials comparing pregnancy or live birth rates after IVF or ICSI for male infertility.

Given the lack of reliable evidence, the Third Eivar Workshop Group advocated the use of a sibling oocyte technique (i.e. half of the oocytes are inseminated by conventional IVF and half by ICSI) if the likelihood of spontaneous fertilization is uncertain. However, the group acknowledged that use of this technique depends on the retrieval of a sufficient number of oocytes.

Bilateral tubal occlusion

The use of IVF and ICSI in a sibling oocyte design study of couples with tubal infertility and normal semen showed similar mean fertilization rates of 53% and 62%, respectively (Staessen et al., 1999). Moreover, no significant benefit in clinical pregnancy or live birth rates was demonstrated in an analysis of data from two pseudo-randomized controlled trials of
couples with tubal infertility who had been treated with ICSI or IVF (Aboulghar et al., 1996; Bukulmez et al., 2000).

Polycystic ovary syndrome

IVF is currently a third-line treatment option, after ovulation induction and laparoscopic ovarian surgery, for women with anovulatory infertility due to polycystic ovary syndrome (PCOS), but it has been suggested that older women may benefit from earlier use of IVF (Thessaloniki ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group, 2008). Analysis of baseline characteristics may enable identification of a subgroup of women who would benefit from a tailored treatment approach (Thessaloniki ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group, 2008). However, further work is required to determine the optimal gonadotrophin stimulation protocol for women with PCOS and the maternal age threshold for deviation from the standard treatment algorithm.

Unexplained infertility

IUI is a commonly used treatment strategy for couples with unexplained infertility (Verhulst et al., 2006). Data have indicated a significantly higher live birth rate with IUI plus ovarian stimulation than with IUI alone for unexplained infertility (OR 2.07, 95% CI 1.22–3.50) (Verhulst et al., 2006).

The use of ovarian stimulation in combination with IUI is a controversial and heavily debated topic (Fauser et al., 2005; Goverde et al., 2005; van Rumste et al., 2006). Biologically, the use of ovarian stimulation would be expected to increase the likelihood of a multiple pregnancy because the development of multiple dominant follicles and ovulation of multiple oocytes is the aim of this intervention. However, published literature contains conflicting reports, which may be related to failure to induce multifollicular development or cycle cancellation after the detection of three or more pre-ovulatory follicles in half of the patients in some studies (Goverde et al., 2005; Verhulst et al., 2006; van Rumste et al., 2006). Nonetheless, a recent systematic literature review of studies of controlled ovarian stimulation and IUI clearly demonstrated that multiple pregnancy rates correlated positively with the number of pre-ovulatory follicles (van Rumste et al., 2008).

Furthermore, the validity of pregnancy rates per cycle for comparisons of different treatment modalities could be questioned (Fauser et al., 2005) when some interventions (e.g. timed intercourse or natural cycle IUI) may be less costly or time-consuming and associated with less patient discomfort and fewer complications. Regardless of the treatment outcome evaluated, the workshop group believes that singleton pregnancy should be the primary aim of ART treatment and, as such, the benefits of ovarian stimulation in combination with IUI must be weighed carefully against the potential risks of a multiple pregnancy.

No differences in live birth rates were demonstrated with IVF or IUI either with (OR 1.15, 95% CI 0.55–2.4) or without (OR 1.96, 95% CI 0.88–4.4) ovarian stimulation in a Cochrane meta-analysis of clinical trials of unexplained infertility (Pandian et al., 2005). ICSI is associated with a significantly lower rate of complete fertilization failure in cases of unexplained infertility than conventional IVF (0.8% versus 19.2%; P < 0.001) (Jaroudi et al., 2003). However, data from three randomized controlled trials indicate that clinical pregnancy rates are similar following IVF or ICSI: 11% and 28%, respectively, per oocyte retrieval (P = 0.09) (Jaroudi et al., 2003); 32% and 38% (RR 0.83, 95% CI 0.48–1.45) (Bhattacharya et al., 2001); and 50% for both [not significant (P-value not provided)] (Foong et al., 2006).

In summary, evaluation of disease characteristics can help clinicians to select the most appropriate active therapeutic option, thus allowing basic tailoring of ART treatment to the individual. However, there is little convincing evidence on which to base treatment strategies for the majority of infertile couples (Table I). More high-quality data on the relative superiority of each treatment option, and associated adverse events, are needed to guide ART management decisions. Comparison of well-defined outcomes, such as (singleton or term) live births per started treatment strategy rather than per treatment cycle, should also be considered. Moreover, outcomes should be viewed in the wider context of time taken to achieve pregnancy, patient stress, likelihood of complications and overall costs of the intervention and monitoring.

**Individualized ovarian stimulation**

**Ovarian stimulation protocols**

The central paradigm of all ovarian stimulation protocols is to maximize the beneficial effects of treatment (relating to high-quality oocyte yield) while minimizing the potential risks associated with

---

**Table I Summary of available evidence on which to base active ART management decisions**

<table>
<thead>
<tr>
<th>Indication</th>
<th>IUI versus IVF</th>
<th>Conventional IVF versus ICSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild-to-moderate male infertility</td>
<td>No difference in clinical pregnancy rates (Crosignani and Walters, 1994; Goverde et al., 2000)</td>
<td>ICSI increases fertilization rates (RR 2.2, 95% CI 1.6–3.0) (Tournaye et al., 2002)</td>
</tr>
<tr>
<td>Tubal infertility</td>
<td>Not applicable</td>
<td>Decreases fertilization failure (Tournaye et al., 2002)</td>
</tr>
<tr>
<td>Unexplained infertility</td>
<td>No difference in live birth rates (Pandian et al., 2005)</td>
<td>Decreases fertilization failure (Jaroudi et al., 2003)</td>
</tr>
</tbody>
</table>

ART, assisted reproductive technology; ICSI, intracytoplasmic sperm injection; IUI, intrauterine insemination; IVF, in vitro fertilization.

---
OHSS and multiple pregnancy (Fauser et al., 2008). The amount of exogenous FSH required to induce follicle development is related to the so-called FSH threshold and varies widely among women (Brown, 1978; Schoemaker et al., 1993; Fauser and Van Heusden, 1997). Conventional daily doses of FSH in ART treatment protocols range from 150 to 225 IU, but close monitoring and dose adjustment is required because of the considerable inter-individual variability in ovarian response (Fauser et al., 2008).

A requirement for elevated doses of FSH to induce multifollicular development and increase the oocyte yield may indicate that a patient is at the extreme of ovarian reserve (Tarlatzis et al., 2003). Low mitochondrial DNA content has been reported to occur more frequently in women with ovarian insufficiency than in those with normal ovarian function (May-Panloup et al., 2005). The consequences of impaired mitochondrial function in human oocytes are currently unknown, but animal studies have shown that replacement of ooplasm in ooplasm-deficient eggs will restore their developmental capacity (Levron et al., 1996). The observed high rate of pregnancy loss in women with reduced ovarian reserve (evidenced by elevated basal FSH levels) indirectly supports the notion of a qualitative reduction in oocyte quality (Levi et al., 2001).

The efficacy and safety profile of standard FSH stimulation protocols vary according to individual patient characteristics. Four factors (FSH, body mass index, female age and AFC) have been identified to be predictive of ovarian response to FSH stimulation in women aged <35 years undergoing ART (Howles et al., 2006). These factors have been combined into a model that predicts the optimum starting dose of recombinant human (rh) FSH for individual patients (Howles et al., 2006; Olivennes et al., 2009). Similar factors were previously shown to predict the dose of exogenous FSH required to induce ovulation in normogonadotrophic anovulatory patients (Imani et al., 2002). It is currently unclear from efficacy, safety, tolerability or health economic perspectives whether mild or maximal ovarian stimulation (with cryopreservation of supernumerary embryos) is most beneficial (Fauser et al., 2005). Mild stimulation protocols [comprising a gonadotrophin-releasing hormone (GnRH) antagonist for pituitary suppression and low-dose FSH stimulation in the mid-to-late follicular phase] may offer lower multiple pregnancy rates and lower costs than standard ovarian stimulation, despite similar overall treatment efficacy (Heijnen et al., 2007). Accordingly, research efforts are ongoing to optimize the efficacy and safety profiles of stimulation protocols.

**Evaluation of ovarian response to stimulation**

There are no universally accepted definitions of normal, poor or excessive responses to ovarian stimulation and the inconsistent definitions used in different studies have hampered efforts to compare treatment outcomes (Tarlatzis et al., 2003; Kailasam et al., 2004). Many definitions relate to retrieved oocyte number, which is a crude marker of response when used in isolation. Ovarian response must be evaluated with reference to the stimulation protocol used, including the daily and/or total dose of FSH administered and co-treatment regimens (Kailasam et al., 2004; Verberg et al., 2009a). Indeed, the retrieval of a small number of oocytes is the very aim of mild stimulation protocols, although this would represent a poor response following conventional ovarian stimulation (Verberg et al., 2009b). Furthermore, it has been suggested that ovarian response should be considered in the context of an individual’s expected response to gonadotrophin stimulation (Klinkert et al., 2004).

**Excessive ovarian response**

The most serious iatrogenic complication of multifollicular ovarian stimulation is severe OHSS, which is believed to be triggered by human chorionic gonadotrophin (hCG) (Aboulghar and Mansour, 2003). This is a life-threatening condition, in which increased capillary permeability results in haemoconcentration and hypovolaemia (Papanikolaou et al., 2005).

There is some evidence that the incidence of early OHSS is limited by the use of a GnRH antagonist (rather than a GnRH agonist) for pituitary down-regulation (Al-Inany et al., 2006; Heijnen et al., 2007), low-dose and mild FSH-stimulation protocols (Heijnen et al., 2007), a GnRH agonist (Griesinger et al., 2007b; Engmann et al., 2008) or low doses of hCG (Nargund et al., 2007) to trigger final oocyte maturation. Although coasting (withholding of gonadotrophs) is often used to prevent OHSS (Mansour et al., 2005; Nardo et al., 2006), a systematic review concluded that there is a shortage of randomized controlled trials investigating such an approach for the prevention of OHSS (D’Angelo and Amso, 2002). The transfer of frozen–thawed embryos in natural cycles may be used instead to limit the incidence of late OHSS (Mathur et al., 2007). However, there are limited data to support the cryopreservation of all embryos (D’Angelo and Amso, 2007). Large randomized trials are required to compare the risks and benefits of different ART treatment strategies for women at risk of OHSS (Aboulghar and Mansour, 2003).

**Poor ovarian response**

Kailasam et al. (2004) attempted to define a poor ovarian response in a retrospective study of IVF cycles in women aged under 40 years. Based on this limited evidence and practical considerations, the workshop group proposed a definition of poor response to stimulation as the retrieval of fewer than four oocytes (or cycle cancellation following the development of fewer than three follicles) in response to an ovarian stimulation protocol of 225 IU FSH per day (which represents maximal ovarian stimulation) (Hoomans et al., 1999; Harrison et al., 2001; Latin-American Puregon IVF Study Group, 2001; Out et al., 2001; Yong et al., 2003; Out et al., 2004). Parameters of oocyte maturity are not included in this definition.

Signs of diminished ovarian reserve may include age older than 40 years, the presence of fewer than five antral follicles of 2–5 mm in diameter on TVS prior to treatment or an elevated basal FSH level (Klinkert et al., 2004; Klinkert et al., 2005). A poor response to gonadotrophin stimulation would be unexpected in patients who have no signs of decreased ovarian reserve (Klinkert et al., 2004). A retrospective cohort study showed higher cumulative pregnancy rates (after a maximum of three cycles) among patients who had experienced an unexpected versus expected poor response in their first IVF treatment cycle (37–47% versus 16–19%) (Klinkert et al., 2004). Thus, patients who have an unexpected poor response to stimulation still have a reasonable chance of pregnancy in subsequent cycles (Klinkert et al., 2004; Hendriks et al., 2008).
There is little convincing evidence to indicate the most appropriate treatment regimen for patients with an expected or proven poor response to ovarian stimulation (Shanbhag et al., 2007). Indeed, prospective, randomized studies have shown little benefit from daily doses of rhFSH of ≥ 300 IU to induce follicular development (Tarlatzis et al., 2003; Kailasam et al., 2004). Moreover, although evidence is conflicting, high doses of FSH may stimulate the recruitment of immature or chromosomally abnormal oocytes or have a detrimental effect on the endometrium (Check et al., 1999; Katz-Jaffe et al., 2005; Kok et al., 2006; Baart et al., 2007). In addition, chromosomal aneuploidy has been observed in embryos from unstimulated IVF cycles in young women (Verpoest et al., 2008).

Future elucidation of relevant genetic markers may allow the identification, prior to stimulation, of patients who are at risk of a poor response or OHSS (Fauser et al., 2008).

**Ovarian resistance to FSH**

Early studies identified a subgroup of normogonadotrophic patients who have normal estimated ovarian reserves but suboptimal responses to FSH stimulation (De Placido et al., 2001, 2004, 2005; Mochtar et al., 2007). Such women express ovarian resistance to FSH but seem to be distinct from classical poor responders because some investigators suggest that luteinizing hormone (LH) supplementation improves their ART treatment outcomes (Alviggi et al., 2006). The incidence of ovarian resistance to FSH is poorly defined in the published literature, but it is believed to affect 10–15% of women with normal ovarian reserve who undergo standard cycles of ART (Alviggi et al., 2006). This subgroup can be identified by estradiol levels of < 180 pg/ml or the absence of follicles of > 10 mm in diameter on Day 8 of a 150–300 IU FSH per day stimulation cycle following GnRH agonist long-protocol pituitary down-regulation (Alviggi et al., 2006).

It is believed that androgens, which are endogenously produced in response to LH stimulation, are involved in sensitizing small antral follicles to FSH (Durnerin et al., 2008). Profound suppression of endogenous LH by GnRH agonists combined with the use of rhFSH may cause LH activity to fall below a hypothetical threshold value in some women. An observational trial by Alviggi et al. (2006) showed that a common polymorphism of LH (v-LH) was present in a higher proportion of patients who required > 3500 IU compared with < 3500 IU FSH for follicular maturation (35% versus 2.6%). v-LH has a short half-life and may be ineffective in supporting FSH-stimulated multifollicular growth.

Despite these data, a meta-analysis showed no significant difference in live birth rates with or without LH supplementation of FSH stimulation in a non-selected patient population (OR 0.92, 95% CI 0.65–1.31); subgroup analyses also produced similar findings (Kolibianakis et al., 2006). However, some of the studies reported above were excluded from this meta-analysis because of co-treatment with variable doses of FSH. Indeed, in a recent Cochrane systematic review, a subanalysis of data from three clinical trials showed a significantly higher pooled estimate of ongoing pregnancy in poor responders who had received co-treatment with rhLH than FSH alone (OR 1.85, 95% CI 1.10–3.11) (Mochtar et al., 2007). Additional data are awaited, and pharmacogenomic studies may provide evidence of associated biological pathways.

**Embryo assessment**

The use of embryo morphology scoring was prospectively evaluated by Holte et al. The group assessed the ability of five real-time visual embryo scoring variables to predict implantation of an individual embryo after double-embryo, Day 2 transfer (Holte et al., 2007). Cleavage stage, variation in blastomere size and the presence of multinucleated cells were found to be independent markers of embryo competence following the application of a conditional multiple-regression model.

An increased understanding of the metabolic needs of embryos and improvements in cell culture media have enabled prolonged in vitro culture of embryos (Fig. 2) (Papanikolaou et al., 2008). Extension of the in vitro culture period now allows the transfer of Day 5 blastocyst-stage rather than early cleavage-stage embryos (Blake et al., 2007). A recent meta-analysis showed significantly higher rates of clinical pregnancy (OR 1.27, 95% CI 1.03–1.55; P = 0.02) and live births (OR 1.39, 95% CI 1.10–1.76; P = 0.005) following the transfer of blastocyst-stage compared with cleavage-stage embryos (Papanikolaou et al., 2008).

Given the speed at which laboratory techniques have evolved, it is essential to compare procedures and outcomes to elucidate the best practices for maximal embryo viability.

The scientific rationale for blastocyst transfer is to increase implantation rates by improving uterine and embryonic synchrony and allowing self-selection of embryos with greater implantation potential (Wilson et al., 2002; Blake et al., 2007). However, there is little level I evidence on which to base recommendations for morphology screening or blastocyst selection.

Greater standardization of blastocyst morphology scales is required to allow comparison of study findings. In addition, validated, objective measurements of blastocyst morphology and growth kinetics throughout the culture period are required to increase the reproducibility of results. Operator-independent computer-processing programmes have shown promising ability to objectively assess blastocyst morphology (Lemmen et al., 2008), but require further validation prior to use in routine practice.

**Selection of culture conditions**

Individual laboratory culture strategies are complex and highly varied, and there are few data on which to base the selection of a specific culture strategy. For example, each culture fluid contains up to 80 components that will influence embryo development and assessments. Greater standardization of culture conditions is required to permit accurate comparisons of embryo and/or blastocyst quality-assessment tools.

**Chromosomal evaluation**

There is a high incidence of chromosomal abnormalities in human embryos cultured in vitro (Munne, 2006) and the incidence of aneuploidy increases with maternal age (Munne et al., 1995; Marquez et al., 2000). Theoretically, identification of chromosomal abnormalities would permit selection of the healthiest embryos, and thus increase live birth rates (Gianaroli et al., 1997; Wells and Delhanty, 2000). Unfortunately, embryo morphology correlates only partially with its chromosomal status (Wells and Delhanty, 2000).
Preimplantation genetic screening (PGS) for aneuploidy can currently be performed in up to 15 chromosomes using fluorescence in situ hybridization (Munne, 2006). PGS requires removal and analysis of a single blastomere from a Day-3 embryo (Gianaroli et al., 1997; Wells and Delhanty, 2000). Embryos with a normal genetic constitution are subsequently selected for transfer and those with an abnormal number of chromosomes are discarded (Mastenbroek et al., 2007).

The success of PGS depends on the assumption that the chromosomal constitution of a single blastomere is representative of the entire embryo. However, mosaicism is now known to occur frequently in early cleavage-stage embryos (Baart et al., 2006; Coulam et al., 2007; Frumkin et al., 2008). The chromosomal constitution of a preimplantation embryo may evolve during early cleavages, allowing some mosaic embryos to develop normally (Baart et al., 2006; Frumkin et al., 2008). Self-corrective mechanisms may include the growth advantage of normal cells within a mosaic embryo (Bielanska et al., 2002) or, in cases of trisomy, the active loss of the additional chromosome (Munne et al., 2005; Frumkin et al., 2008).

Despite the promising results of early observational studies (Gianaroli et al., 1999; Munne et al., 1999; Munne et al., 2003; Montag et al., 2004), recent randomized controlled trials have shown no improvement in pregnancy outcomes after PGS (Staessen et al., 2004; Twisk et al., 2006; Mastenbroek et al., 2007; Hardarson et al., 2008; Staessen et al., 2008; Fauser, 2008). Indeed, the use of PGS may even reduce resulting pregnancy and live birth rates for couples with advanced maternal age (Mastenbroek et al., 2008; Fauser, 2008).

In short, although evidence is conflicting (Goossens et al., 2008), there is no good-quality evidence to support a benefit of routine PGS in the prediction of pregnancy outcomes (Staessen et al., 2004; Baart et al., 2006; Twisk et al., 2006; Coulam et al., 2007). Thus, current methods of PGS for aneuploidy have no place in current ART treatment protocols (Twisk et al., 2006; Mastenbroek et al., 2007; Fauser, 2008; Practice Committee of Society for Assisted Reproductive Technology and Practice Committee of American Society for Reproductive Medicine, 2008b).

Cryopreservation techniques

Cryopreservation of gametes, embryos and blastocysts is an essential component of modern ART (Yousry et al., 2008). Successful cryopreservation maximizes cumulative pregnancy rates per oocyte retrieval by storing unused embryos for future use (Yousry et al., 2008) and allows delayed embryo transfer during a natural menstrual cycle (Loutradi et al., 2008). Two cryopreservation techniques are available: the conventional slow cooling method and the recently introduced rapid procedure known as ‘vitrification’ (Yousry et al., 2008).

Vitrification results in significantly improved survival of frozen–thawed cleavage stage blastocysts and pregnancy rates, compared with standard cryopreservation (Fig. 3) (Loutradi et al., 2008; Yousry et al., 2008). However, there are concerns regarding the high concentrations of potentially cytotoxic cryoprotectants used during the vitrification process (Loutradi et al., 2008).

Chromosomal abnormalities have been reported in rapidly frozen mouse embryos (Shaw et al., 1991). Thus, there is a potential risk of serious malformations in human embryos following cryopreservation. Although no adverse effects of the transfer of cryopreserved embryos on IVF outcomes, including malformations, were shown in two European registry studies (Kallen et al., 2005; Pinborg et al., 2008), more high-quality studies are required to assess long-term
safety, given the recent and rapid development of cryopreservation techniques (Shaw et al., 1991).

Frozen–thawed embryo or blastocyst transfer schedule

Transfer of frozen–thawed embryos or blastocysts can be performed during spontaneous ovulatory cycles, cycles of ovulation induction, cycles in which the endometrium is artificially prepared using estrogen and progesterone, and all with or without the use of a GnRH agonist (Ghobara and Vandekerckhove, 2008). Despite these numerous options, a recent Cochrane review found insufficient evidence to support the use of a particular transfer schedule (Ghobara and Vandekerckhove, 2008).

Laboratory standards and procedures

High laboratory standards and adherence to correct procedures are critical to the outcome of ART. Unfortunately, there is currently little good-quality evidence to support the use of specific ART-related laboratory processes, including culture strategy and selection of blastocysts for transfer. Furthermore, there is considerable variation in quality and outcomes between IVF clinics and laboratories. This is reflected in the considerable variation in cycle outcomes recorded at individual UK clinics despite a country-wide policy of a maximum of two embryos transferred (Human Fertilization and Embryology Authority, 2008).

The observed inter-centre variability hampers agreement on, and uptake of, best clinical and laboratory practices. In an attempt to standardize IVF laboratory techniques, the ESHRE has developed guidelines for good practice that require robust quality-management programmes for each ART centre (Magli et al., 2008). Well-powered studies to investigate ART-related laboratory practices and monitor adherence to the recently published guidelines are warranted, and we encourage professional bodies to promote such research efforts.

Single embryo transfer

Despite recommendations to limit the number of embryos transferred (The ESHRE Capri Workshop Group, 2000; Land and Evers, 2003; Wright et al., 2003; Practice Committee of Society for Assisted Reproductive Technology and Practice Committee of American Society for Reproductive Medicine, 2008a), at least 20–30% of IVF-related pregnancies are twin or higher-order multiple gestations (Reddy et al., 2007; Andersen et al., 2008). Furthermore, approximately half of all babies born from IVF globally are from multiple pregnancies (Adamson et al., 2006).

Multiple pregnancy and birth is associated with a high incidence of maternal and neonatal complications (Pandian et al., 2004; Wood, 2008) and, thus, considerable fetal morbidity and costs (ESHRE Campus Course Report, 2001; Land and Evers, 2003). Indeed, whenever possible, elective single embryo transfer (eSET) is recommended for young patients (aged <35 years) (Practice Committee of Society for Assisted Reproductive Technology and Practice Committee of American Society for Reproductive Medicine, 2008a).

Randomized controlled trials have clearly shown that when only fresh transfer cycles are evaluated, double embryo transfer (DET) leads to higher pregnancy and live birth rates than eSET (Gerris et al., 1999; Martikainen et al., 2001; Thurin et al., 2004; Lukassen et al., 2005; van Montfoort et al., 2006). However, comparable rates of cumulative clinical pregnancy (52.6% versus 47.9%; $P = 0.24$) and live births (42.9% versus 38.8%; $P = 0.30$) have been demonstrated following DET or eSET, respectively, when a single frozen–thawed embryo transfer is also considered as part of the same stimulation cycle (Thurin et al., 2004). Unsurprisingly, a recent Cochrane meta-analysis demonstrated a significantly lower multiple pregnancy rate in women who underwent SET than DET (OR 9.97, 95% CI 2.61–38.19; $P = 0.0008$) (Pandian et al., 2004). Health economic data also support the superior cost-efficacy of SET over DET when the number of deliveries with at least one live-born child and all complications are considered (Kjellberg et al., 2006; Polinder et al., 2008; Wood, 2008).

The group proposed that an acceptable rate of ART-related multiple pregnancies would be up to 10%. However, the worldwide

<table>
<thead>
<tr>
<th>Study</th>
<th>Vitrification n/N</th>
<th>Slow freezing n/N</th>
<th>OR (random) 95% CI</th>
<th>Weight %</th>
<th>OR (random) 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rama Raju</td>
<td>121/127</td>
<td>72/120</td>
<td>34.00</td>
<td>13.44</td>
<td>[5.48, 32.98]</td>
</tr>
<tr>
<td>Zheng</td>
<td>46/49</td>
<td>8/52</td>
<td>28.71</td>
<td>84.33</td>
<td>[21.01, 338.51]</td>
</tr>
<tr>
<td>Kuwayama</td>
<td>879/997</td>
<td>857/942</td>
<td>37.28</td>
<td>4.84</td>
<td>[2.89, 6.12]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>1073</td>
<td>1114</td>
<td>100.00</td>
<td>15.57</td>
<td>[3.68, 65.82]</td>
</tr>
</tbody>
</table>

CI, confidence interval; OR, odds ratio. Total events: 1064 (vitrification), 937 (slow freezing). Test for heterogeneity: $\chi^2 = 15.94$, df = 2 ($P = 0.001$). Test for overall effect: $Z = 3.73$ ($P = 0.0002$). Reprinted from Loutradi et al. (2008) with permission from Elsevier.
incidence of ART-related multiple births remains considerably higher (Reddy et al., 2007; Andersen et al., 2008). The observed resistance to universal uptake of eSET in some societies is related to its perceived low efficacy; factors including advanced maternal age and few or poor-quality oocytes retrieved are cited frequently to justify the transfer of multiple embryos (ESHRE Campus Course Report, 2001; van Montfoort et al., 2006; Practice Committee of Society for Assisted Reproductive Technology and Practice Committee of American Society for Reproductive Medicine, 2008a). This is confounded by insufficient awareness of the risks and costs associated with multiple pregnancy among the general public and policy makers, the inability to select the best-quality embryo, suboptimal cryopreservation processes, constrictive cryopreservation legislation, the costs related to repeated treatment cycles (particularly in low- and middle-income countries) and competition between ART centres based on pregnancy/birth rates per cycle (Serour et al., 1991; ESHRE Campus Course Report, 2001; Adashi et al., 2003; Fauser et al., 2005; Karlstrom and Bergh, 2007).

**Evaluation of ART treatment outcomes**

The single most relevant measure of success in ART is a controversial and heavily debated topic (Land and Evers, 2003; Dickey et al., 2004; Griesinger et al., 2004; Heijnen et al., 2004; Pinborg et al., 2004). Previously proposed criteria to evaluate successful ART treatment outcomes include live birth rate per ovarian stimulation started (Griesinger et al., 2004), healthy live birth rate per treatment cycle (Dickey et al., 2004), singleton and multiple live birth rates per started treatment cycle (Vayena et al., 2001) or term live birth per started treatment strategy, which may include multiple cycles (Heijnen et al., 2004). Because of the lack of a consistent definition of ART success, national criteria for evaluation of treatment outcomes tend to reflect the local economic and legal frameworks.

The Third EVAR Workshop Group concurred with the ESHRE recommendations and believes that the birth of a single healthy child should be the aim of ART (Land and Evers, 2003). Therefore, the group advocated the use of singleton delivery rates as the gold standard expression of ART treatment outcome. Nonetheless, cumulative delivery rates per retrieved oocyte cohort using fresh and frozen embryos, or per treatment strategy, must also be considered. The group also believed that all comparisons of treatment modalities must incorporate efficacy and safety data, in addition to health-economic evaluations.

**The future of ART**

As described, the Third EVAR Workshop Group believes that optimization of singleton delivery rates should be the common aim of all ART clinicians. Reproductive medicine specialists have a responsibility to educate policymakers and the wider society on the risks of multiple pregnancies and births. Furthermore, the group calls on national and international professional bodies to issue guidelines promoting a responsible attitude to eSET, and to help raise awareness of the greater cost-efficacy of SET compared with DET among healthcare providers and policymakers.

---

**Figure 4** New techniques for assessment of oocyte and embryo quality.

CGH, comparative genomic hybridization; FISH, fluorescence in situ hybridization.
Future refinements in cryopreservation and vitrification techniques are expected to have a huge impact on ovarian stimulation protocols and increase the acceptance and uptake of SET around the world. Improvements in cryopreservation outcomes may also encourage the use of thawed embryos in subsequent natural cycles. This process may help to avoid the consequences of gonadotrophin stimulation on endometrium and late-onset OHSS.

New techniques in assessing oocyte and embryo quality could also improve pregnancy and delivery rates per embryo transfer, thus encouraging greater uptake of SET (Fig. 4). This may be achieved by studying oocyte zona birefringence (Montag et al., 2008), gene expression profiling of oocyte cumulus cells (McKenzie et al., 2004; Assou et al., 2006; Feuerstein et al., 2007), evaluation of spent culture fluid by proteomic analysis (Katz-Jaffe et al., 2006) or metabolic profiling using near-infrared spectroscopy (McKenzie et al., 2004; Vergouw et al., 2008). Good-quality studies of these techniques will help evaluate their potential clinical application and contribute to our understanding of basic oocyte and embryo biology, and the effects of iatrogenic ovarian stimulation.

The use of GnRH agonists to trigger final oocyte maturation during GnRH antagonist cycles is another promising approach to ART (Griesinger et al., 2007a). Results of a small, observational, proof-of-concept study suggest that GnRH agonist triggering (in combination with elective cryopreservation of two pronucleate oocytes) leads to acceptable cumulative pregnancy rates while reducing the risk of moderate-to-severe OHSS. Further investigation and validation of these new ART treatment strategies in prospective studies are eagerly awaited. Accurate identification of (genetic) markers of ovarian response to gonadotrophin stimulation would strengthen the development of predictive models of response and may enable the use of truly individualized ART stimulation treatment protocols (Fauser et al., 2008).

Conclusions

In conclusion, members of the Third EVAR Workshop Group agreed that SET should be the primary aim of many ART treatment cycles and supported the use of singleton live birth rate as the reported outcome measure from clinical trials and routine practice. However, the group acknowledged that improved cryopreservation techniques are required to further increase the global uptake of SET. Within this standard ART framework, adaptation and personalization of therapy may help to optimize efficacy and safety outcomes for individual patients.

Management decisions, including expectant therapy versus IUI, IVF or ICSI, reflect a rudimentary individualization of therapy but are currently based on limited available evidence. Greater quality control and standardization of clinical and laboratory evaluations are needed to optimize ART practices and improve individual patient outcomes. Furthermore, additional well-designed, good-quality studies are required to drive improvements in the diagnosis and management of ART processes in future years.

Acknowledgements

The authors thank Hannah Wills of Caudex Medical (supported by Merck Serono S.A.—Geneva) for her assistance in the preparation of this manuscript.

Funding

The workshop and preparation of the manuscript were sponsored by an unrestricted educational grant from Merck Serono S.A.—Geneva. P.D. has received grant support from Ferrering, Merck Serono S.A., Organon (now part of Schering-Plough) and Besins. B.C.J.M.F. has received fees and grant support from the following companies: Andromed, Ardana, Ferrering, Genovum, Merck Serono S.A.—Geneva, Organon (now part of Schering-Plough), Pantarhei Bioscience, PregLem, Schering, Schering-Plough, Serono (now Merck Serono S.A.—Geneva) and Wyeth.

References


Improving diagnosis and management of infertility


Dickey RP, Sartor BM, Pyrzak R. What is the most relevant standard of success in assisted reproduction? No single outcome measure is satisfactory when evaluating success in assisted reproduction; both twin births and singleton births should be counted as successes. Hum Reprod 2004;19:783–787.


Improving diagnosis and management of infertility

407


Submitted on November 5, 2008; resubmitted on February 19, 2009; accepted on March 18, 2009.