Optimal use of infertility diagnostic tests and treatments

The ESHRE Capri Workshop Group*

The general definition of infertility is a lesser capacity to conceive than the mean capacity of the general population and infertile couples can be characterized in two groups: those unable to conceive without therapy and those who are hypofertile, but conceive without therapy. The initial diagnostic tests for infertility should include a midluteal phase progesterone assay, a semen analysis and a test for tubal patency such as a hysterosalpingogram. Measuring progesterone is the best test for confirming ovulation. To predict ovulation, evaluating the luteinizing hormone (LH) surge is the best single assay while measurement of LH plus preovulatory oestrogen is the best prediction. Today primary investigation of the morphology of the uterus and tubes should be by hysterosalpingography. However, ultrasound, particularly with simple contrast media, is likely to gain in importance. Laparoscopy should be reserved as a further diagnostic procedure or in combination with endoscopic surgery. There are situations in which semen analysis is of utmost importance and of absolute predictive value, namely, in cases of azoospermia. In general semen analysis remains a substantial part of the fertility workup, but any consideration of its predictive value has to be cautious. Performing genetic tests before, during and after assisted reproductive techniques (ART) is an intrinsic part of good clinical practice. These tests allow one to reach a correct diagnosis, to give adequate genetic counselling to the couple and their families in cases such as (i) women with Turner syndrome; (ii) men with 47,XXY; (iii) men or women with structural chromosomal aberration; (iv) men with Yq11 deletion or (v) men with congenital bilateral absence of vas deferens. Patients should, of course, be made aware of the occurrence of de-novo mutations taking place in the testis and in the embryo. Treatment of some causes of infertility are of proven value. For example induction of ovulation. Others are more controversial. Among the many empirical treatments suggested for the treatment of the various form of subfertility, surgical treatment of varicocele in the male, treatment of pelvic endometriosis in the female and the efficacy of the ART strategies offered to the subfertile couple are considered. Many varicocele studies are of poor quality. A few are good, but small in size. They do not show an improvement in pregnancy rates. Therefore, at the moment, there is insufficient scientific evidence for recommending routinely surgical treatment in subfertile and/or oligozoospermic men with a varicocele. Randomized, double-blind controlled trials demonstrated the modest efficacy of endometriosis ablation in increasing the pregnancy rate in infertile women while drugs suppressing ovulation are of no benefit to infertile women with endometriosis. Although the largest body of evidence available suggests that IVF success declines in repeated ART cycles, an accurate estimate of the true success rate in the ‘nth’ cycle of IVF treatment is not possible. Similarly little is still known of the reasons for the overall low continuation rates with IVF treatment.

Critical evaluation of the diagnosis of infertility

The classical criteria that determine the usefulness of any diagnostic test are:

- Sensitivity: to produce few false negatives.
- Specificity: to produce few false positives.
- Invasiveness: harmfulness.
- Complexity and the time required are important, because if a situation is very complex, treatment may be delayed and the waiting time increased for other couples.
- Cost.
- Usefulness: Is there any treatment if the result is positive? Does knowing the result alter the management in any way?
- Positive predictive value. For the infertile couple, wherever possible this should be expressed in terms of pregnancy.
- Negative predictive value. Expressed not only in terms of sensitivity but also in terms of pregnancy.

Testing habits are difficult to erase and technology becomes ever more sophisticated and readily available (e.g. Falloporyscope). Thus, both old and new diagnostic tests must be considered, but to what degree is diagnostic certainty necessary? The science of infertility is uncertain and it is not a life-threatening condition. Testing until uncertainty vanishes may delay treatment (and if the delay is long enough, the female patient will become menopausal).

Clinical evaluation of the infertile couple

The general definition of infertility is a lesser capacity to conceive than the mean capacity of the general population.
The specific definition of infertility is inability of a couple to conceive after 1 year of sexual intercourse without contraception. Fecundability is defined as the rate of conception occurring in a population in a given time period. The monthly conception rate of normal fertile couples is ~20%. In couples where the woman ovulates regularly and has no genital tract abnormalities, and the male partner has a normal semen analysis, the fecundability rate steadily decreases with increasing numbers of cycles in which conception does not occur. In addition fecundability steadily decreases with increasing age of the female partner beyond the age of 31 years (van Noord-Zaadestra et al., 1991). Cigarette smoking and caffeine intake by the female, but not the male partner, also decrease fecundability rates (Hatch and Bracken, 1993; Bolumar et al., 1996; Angood et al., 1998).

Infertile couples can be characterized in two groups, i.e. those unable to conceive without therapy and those who are hypofertile, but who, with time, will probably conceive without therapy. Examples of the former group are women with complete tubal occlusion or anovulation and men who are azoospermic. Examples of the latter group include women with mild endometriosis, men with oligozoospermia and couples with unexplained infertility (ESHRE Capri Workshop, 1996). Therefore, to determine whether an infertile couple has an abnormality known to cause infertility, it is important to perform diagnostic tests to determine if ovulation is occurring, the oviducts are patent and uterine cavity is regular, and the man produces sufficient number of spermatozoa to allow the ovum to be fertilized.

Thus the initial diagnostic tests for infertility should include a midluteal phase progesterone assay, a semen analysis and a test for tubal patency such as a hysterosalpingogram. Visualization of the pelvic cavity by endoscopy is also necessary to determine if there are adnexal adhesions that could interfere with ovum retrieval by the oviducts, as well as to determine if there is pelvic endometriosis, although it has not been definitively demonstrated that mild pelvic endometriosis without tubal adhesions is a cause of infertility (Inoue et al., 1992). Knowing that mild endometriosis is present is unlikely to influence management at this stage. In addition if a Chlamydia trachomatis antibody titre is not elevated (Meikle et al., 1994) and pelvic sonography does not demonstrate the presence of an endometrioma, the incidence of major pelvic abnormalities is found in <5% of the patients.

Numerous other diagnostic tests have been recommended for the evaluation of the infertile couple. These include measurement of thyroid stimulating hormone and prolactin in the ovulatory woman (Glazener et al., 1987a), performance of a luteal phase endometrial biopsy to date the endometrium (Shoupe et al., 1989; Peters et al., 1992) and more recently an estimation of integrins (Lessey et al., 1995; Lessey, 1998), measurement of antisperm antibodies in the semen and circulation (Smarr et al., 1988), and the number of spermatozoa penetrating the zona-free hamster egg, which has a good predictive value (Romano et al., 1998). One criterion to guide the use of such tests is whether the test results might indicate an effective treatment. To determine if any treatment is superior to no treatment, results need to be analysed by life table analysis and comparison with non-treated or historical controls.

Four studies have reported pregnancy rates over time in untreated couples in whom the woman is ovulatory, the oviducts are patent and the semen analysis is normal. After 2 years with no other diagnostic test performed, pregnancy rates ranged from 30% to 70% in the different populations. Epidemiological analyses reach the same conclusion (Basso et al., 1998; Spira, 1998). In view of this, therapy should only be offered if proven to shorten time to conception compared with no treatment; in other words treatment should increase the fecundability rate.

Perceived aetiologies of infertility without a proven causal relation include: (i) mild endometriosis without tubal adhesions; (ii) luteal insufficiency; (iii) antisperm antibodies and (i) hyperprolactinaemia without anovulation or dysovulation. Treatment of these conditions has not been proved to achieve higher conception rate compared to no treatment, making both diagnosis and treatment of these conditions unnecessary.

It can be useful to measure follicle stimulating hormone (FSH) in patients undergoing ovulation induction; to perform pelvic sonography or laparoscopy to decide on management before in-vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI) cycles and to obtain genetic tests when indicated.

### Table 1. Tests for ovulation

<table>
<thead>
<tr>
<th>Prediction Confirmation</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>US (growing follicle)</td>
<td></td>
<td>Oestradiol–LH–mucus</td>
</tr>
<tr>
<td>USb (collapsing follicle)</td>
<td></td>
<td>Progesterone–BBTb</td>
</tr>
</tbody>
</table>

aEssential for luteinized unruptured follicle syndrome.
bHigh incidence of false negative readings.

US = ultrasound; LH = luteinizing hormone; BBT = basal body temperature.

Reliability of the diagnostic tests for ovulation

Detection of ovulation is an important component in the investigation and treatment of infertile couples, as well as the cornerstone of ‘natural’ family planning. In contrast to almost all other species, there are no behavioural or obvious physical signs of impending ovulation in women, although there may be hormonally induced changes in mood and behaviour throughout the ovarian cycle.

Tests can be used to predict or confirm ovulation (Table 1). Prediction is useful in order to time intercourse during the most fertile period in those wishing to become pregnant (Wilcox et al., 1995). Natural family planning requires both the prediction and confirmation of ovulation to define the limits of the fertile period. Several new methods of ovulation prediction and detection have been developed in the last decade. They can all be divided into direct and indirect methods (Vermesh et al., 1987).

**Direct**

Direct methods involve visualization of the process of follicle rupture either through the laparoscope or, more
commonly, by high resolution ultrasound. With transvaginal ultrasound, it is possible to observe collapse of the ovulatory follicle and the appearance of fluid in the pouch of Douglas in the majority of women (Collins et al., 1991; Brännström et al., 1998). These methods are expensive and/or invasive, but are essential for the detection of the luteinized unruptured follicle syndrome (LUF).

**Indirect**

Indirect methods depend on measuring ovarian or pituitary hormones or their biological effects. The increased secretion of oestradiol by the dominant follicle is responsible for inducing a surge of luteinizing hormone (LH) (positive feedback) which usually, but not always, results in follicle rupture beginning 37–38 h later. Thus, all indirect methods of predicting ovulation depend on the measurement of LH and/or oestrogen in blood fluids (blood, urine, saliva), or the effect of oestrogen on cervical mucus. Measurements of oestradiol and LH in blood have high predictive value because the magnitude of the change in concentration of the LH surge is great (O’Connor et al., 1998). All the self-testing kits are based on measurements of LH in urine, which is easy to collect. It is possible to predict the fertile period and detect the LH surge through parallel assay of oestrone glucuronide and LH in the urine. While these urinary methods give a higher incidence of false negative results than radioimmunoassay or immunoradiometry in blood, they are accurate enough to have wide application in self-testing, both for enhancing fertility and for contraception (Robinson et al., 1992; Anderson et al., 1996).

Indirect methods of confirming that ovulation has occurred depend on the measurement of progesterone, its metabolite (pregnanediol) or biological effects [basal body temperature (BBT) or endometrium]. The concentration of progesterone in blood begins to rise within 12 h of the start of the LH surge and has been used as confirmation of ovulation during donor insemination cycles. More commonly, the measurement of progesterone in plasma or serum, or of pregnanediol in urine during the midluteal phase of the cycle, is used to confirm the presence of a corpus luteum. The often imprecise measurement of BBT has largely been abandoned in most practices since the simple chemical assays became available. Biochemical tests are not able to discriminate between ovulation and LUF, the detection of which requires serial ultrasound measurements.

Measuring progesterone is the best test for confirming ovulation. To predict ovulation, evaluating the LH surge is the best single assay, while measurement of LH plus pre-ovulatory oestrogen is the best prediction.

It has been suggested that defective luteal function may be a cause of subfertility. A short luteal phase (<10 days) or inadequate secretion of progesterone (mid luteal progesterone concentration <15 nmol/l) occurs from time to time in fertile women usually in association with defective follicular growth. It is a common occurrence during establishment of ovulatory cycles around menarche, during lactation and in hyperprolactinaemic patients. A short luteal phase can only be diagnosed by identifying the day of ovulation with serial measurements of LH in blood or urine or by ultrasound. Because of the pulsatile nature of progesterone secretion by the corpus luteum, measurement of the concentration of progesterone in at least three samples collected around day 20–24 is necessary to diagnose defective luteal phase (Van Hall and Mastboom, 1979; Balasch et al., 1986; Gibson, 1990).

For couples with unexplained infertility the additional measurement of FSH and PRL could be of importance. Major problems include:

(i) the inability of indirect methods to detect LUF (there is an increased incidence of this syndrome in patients with chronic use of non steroidal anti-inflammatory drugs or in women treated with clomiphene citrate);
(ii) the relatively high incidence of false negative results in methods depending on urine versus blood measurements;
(iii) the difficulty to time ovulation in women with prolonged follicular phases.

Areas requiring more research include:

(i) study of the basic mechanisms involved in ovulation (e.g. role of cytokines);
(ii) development of cheaper, reliable chemical methods for predicting ovulation;
(iii) investigation into the incidence of luteinizing unruptured follicles in spontaneous cycles and in women treated with ovulation-inducing drugs (e.g. anti-oestrogens);
(iv) determination of the minimum height and duration of the LH surge necessary to induce follicle rupture and formation of a normal corpus luteum.

**Optimal evaluation of the morphology of the uterus and the tubes**

Evaluation of the morphology of the uterus and tubes is an integral part of the investigation of the female partner for infertility. The principal options are hysterosalpingography (HSG) and laparoscopy with hydrotubation. The two techniques are complementary, but there is considerable pressure to simplify the investigation and the cost/benefit calculation tends to favour hysterosalpingography. It was demonstrated that laparoscopy made no further contribution to the cumulative conception rate over preliminary HSG (Belisle et al., 1996). HSG has 65% sensitivity and 83% specificity for tubal obstruction (Swart et al., 1995) and has been advocated for predicting fertility (Mol et al., 1997) but it is not a reliable test for a normal tubal patency. Laparoscopy further elucidates pathology but is more invasive and could perhaps be reserved for when it is necessary to elaborate an identified pathology, or to define symptoms. It may then contribute significantly to subsequent management. This is highlighted by the suggestion that ablative therapy of milder degrees of endometriosis increases the subsequent pregnancy rate (Marcoux et al., 1997). However, it is not practical at present for all patients. Sonohysterography and sonosalpingography, elements of the same procedure, are increasingly being used and being demonstrated to be effective, particularly when used with saline or contrast materials (Holz et al., 1997; Vercellini et al., 1997; Hamilton et al., 1998). It seems that the primary investigation of the morphology of the uterus and tubes should be by hysterosalpingography. Cases should be selected by history, physical examination and results of this primary investigation.
Interventions such as tubal surgery, myomectomy or intrauterine manoeuvres may determine the next step and additional procedures may be required to elucidate symptoms other than fertility, such as pain.

However, ultrasound, particularly with simple contra contrast media, is likely to gain in importance. Laparoscopy should probably be reserved as a further diagnostic procedure or in combination with endoscopic surgery.

It remains to be evaluated whether or not sono-HSG is equivalent to HSG. Data obtained with falloposcopy are still not clearly correlated with impaired fecundability but the technique is promising (Heylen et al., 1995; Dechaud et al., 1998). Nevertheless selective salpingography (Gleicher et al., 1993) has been accepted by tubal microsurgeons as a technique to be used before proceeding to tubocornual anastomosis (Dubuisson et al., 1997). Hysteroscopy is only required for confirmation of doubtful uterine pathology and for the relevant therapy.

**Testing the fertilization ability of the ejaculate**

There are situations in which semen analysis is of utmost importance and of absolute predictive value, namely, in cases of azoospermia. Following vasectomy, only azoospermia established in consecutive semen samples will assure complete contraceptive protection; in trials for male hormonal contraception, azoospermia diagnosed by semen analysis has to be reached to ensure contraceptive effectiveness of the treatment (Nieschlag and Behre, 1998). Moreover, in the infertility work-up, azoospermia, repeatedly confirmed, is a clear predictor of absolute infertility.

Aside from azoospermia, multiple studies have demonstrated the degree to which sperm parameters may be of predictive value (ESHRE Capri Working Group). For example even if only a proportion of fathers whose paternity derives from natural conception have ‘normal’ semen values, as proposed by the WHO Manual (Cooper et al., 1991), this does not necessarily imply that these measurements are useless, it may rather mean that ‘normality’ requires redefinition (Coetzee et al., 1998). It is well documented that patients with secondary hypogonadism treated with gonadotrophins or gonadotrophin releasing hormone may induce pregnancies despite very low sperm counts (Büchter et al., 1998). Nevertheless, in the general infertile population chances of inducing pregnancy increase with the number of motile spermatozoa, provided the female partner has no reproductive dysfunction (Glazener et al., 1987b; Bostofte et al., 1990, 1993; Kjaergaard et al., 1990). Thus sperm parameters should be considered of relative predictive value when advising a couple on which therapeutic steps to take (Table II).

Several marker substances extend the diagnostic range of semen analysis. Azoospermia in combination with low α-glucosidase levels (and normal FSH) is indicative of obstruction at the epididymal level or further downstream (Cooper et al., 1990). Low fructose is a sign of seminal vesicle dysfunction, which can be further explored by transrectal ultrasonography. Sperm surface antibodies of the IgG or IgA class in high concentrations are strongly negatively correlated with chances for spontaneous pregnancy (Abshagen et al., 1998). Testing for the acrosome reaction appears to be a logical diagnostic step in the cascade of events leading to fertilization, but its real value still awaits further investigation (Liu and Baker, 1996).

Whereas semen analysis is required to select those infertile men who may benefit most from ICSI, i.e. those with severe oligoasthenoteratozoospermia, conventional semen parameters from these men have failed totally to predict outcome of ICSI (Nagy et al., 1995; Koppers et al., 1998), even when the ‘strict’ criteria for morphology assessment are applied (Svalander et al., 1996; Devroey et al., 1998; Merca et al., 1998). The only important parameter for ICSI success appears to be sperm vitality. The diagnostic value of sperm head DNA condensation or DNA fragmentation (Lopes et al., 1998) remains to be established. It is becoming increasingly clear that spermatozoa are not just fertilizing ‘agents’ carrying chromosomes into an egg, but contain much prognostic information for the fate of the early embryo (Barratt and St John, 1998; Kramer and Krawetz, 1998). Testing for these factors will become of increasing importance in the future. A major problem concerning sperm diagnosis for ICSI is the fact that so far no diagnostic test (except microscopic inspection) can be used on those spermatozoa used for injection into the oocyte. Therefore, the establishment of non-destructive sperm tests is required.

In conclusion, semen analysis remains a substantial part of the fertility work-up, but any consideration of its predictive value has to be cautious.

**Genetic assessment of infertility**

Genetic aberrations have been observed in infertile women and men. These genetic anomalies include:

(i) numerical chromosomal anomalies [e.g. Turner syndrome (45,X) or Klinefelter syndrome (47,XXY)];

(ii) structural chromosomal aberrations;

(iii) microdeletions on the long arm of the Y chromosome; and

(iv) gene mutations in several monogenic conditions (Mak and Jarvi, 1996; Van Assche et al., 1996; Meschede and Horst, 1997; Vogt, 1997; Chandlely, 1998).

Genetic tests are part of the diagnostic work-up of infertile individuals. This approach is different from the genetic screening, which can be carried out when a particular gene mutation is elevated in a population, such as screening for Tay–Sachs mutations in Ashkenazi Jews.

**Genetic testing before assisted reproduction including ICSI**

A careful family and personal history should be taken, because conditions such as recurrent abortions, the birth of a child with multiple congenital anomalies, siblings also presenting with infertility, may indicate abnormalities in karyotype or monogenic diseases (Chandlely, 1998). A careful physical examination may allow the diagnosis of genetic diseases or reveal symptoms pointing to transmissible diseases. Further genetic tests will confirm the diagnosis of a genetic disorder in infertile individuals.

A karyotype should be performed for infertile women suspected of having chromosomal aberrations (e.g. premature
menopause) or in women with habitual abortion. Other genetic tests may also be indicated according to clinical history and findings. A karyotype is indicated for men with severely impaired semen, including non-obstructive azoospermia (Van Assche et al., 1996; Chandley, 1998); a search for AZFa,b,c deletions is also indicated for these men (Vogt, 1997). In couples with congenital bilateral absence of vas deferens (CBAVD) both partners of the couple should be screened for presence of mutations in the cystic fibrosis transmembrane regulator (CFTR) gene (Lissens et al., 1996). Because the spectrum of CFTR mutation is different between populations the ethnic background of the couple should be considered (De Braekeleer and Ferec, 1996; Kanavakis et al., 1998). Other genetic tests may be indicated by history or clinical findings. For research purposes more complex tests such as meiotic studies of testicular biopsies or chromosomal studies of spermatozoa may be carried out in specialized centres.

**Genetic testing during assisted reproduction including ICSI**

Preimplantation genetic diagnosis (PGD) can be carried out during the assisted reproduction procedure. The most relevant example within infertile couples is PGD for cystic fibrosis (CF) in couples where the man has CBAVD due to one or two mutations in the CFTR gene and the wife happens to be a carrier. If PGD cannot be offered, conventional prenatal diagnosis has to be carried out. PGD is also offered to most fertile couples at high risk for severe monogenic conditions and to some couples carrying structural chromosomal aberrations (Liebaers et al., 1998).

It remains a matter of debate whether PGD should also be offered to women over 35 years of age having assisted reproduction, with the aim of avoiding the birth of children with trisomy 13, 18, 21 or a sex chromosomal aberration or to screen the embryos for aneuploidy in order to increase the success rate of assisted reproduction (Gianaroli et al., 1997; Verlinsky et al., 1998).

**Genetic testing after assisted reproduction including ICSI**

The current data indicate that there is a slight but significant increase in de-novo chromosomal aberrations after ICSI compared to the general population: 28 chromosomal aberrations were detected in 1082 fetal karyotypes (Bonduelle et al., 1998). Ten of these were inherited mainly from the father and only one of these was unbalanced and led to termination of pregnancy. The other 18 were either autosomal aberrations or trisomies ($n = 9$) or sex chromosome aberrations ($n = 9$). This information should be used to inform those couples who decide to be treated and to offer them the possibility of prenatal diagnosis.

Pregnancy and newborn follow-up has shown that, after IVF with ICSI, the incidence of major congenital malformations in 1966 live born infants (2.3%) is not higher than that of general population surveys (Tarlatzis and Bili, 1998). Further studies are needed, especially after ICSI with epididymal or testicular spermatozoa and after replacement of cryopreserved ICSI embryos.

In conclusion, performing genetic tests before, during and after assisted reproduction is an intrinsic part of good clinical practice. These tests allow one to reach a correct diagnosis, to give adequate genetic counselling to couples and their families in cases such as (i) women with Turner syndrome; (ii) men with 47,XXY; (iii) men or women with structural chromosomal aberration; (iv) men with Yq11 deletion; or (v) men with CBAVD.

Patients should, of course, be made aware of the occurrence of de-novo mutations taking place in the testis and in the embryo.

**Critical evaluation of treatments offered to the infertile couple**

Treatment of some causes of infertility are of proven value, e.g. induction of ovulation. Others are more controversial.

Among the many empirical treatments suggested for the treatment of the various forms of subfertility, surgical treatment of varicocele in the male, treatment of pelvic endometriosis in the female and the assisted reproduction strategies offered to the subfertile couple will be considered in the following paragraphs.

**Treatment for varicocele**

Does varicocele cause subfertility? Does subfertility cause varicocele? Is there a common denominator for the failure of the genital vasculogenic and spermatogenic system? Arguments in favour of a causal relationship between varicocele and subfertility can be found in some studies reporting a high frequency of occurrence of abnormal semen analysis results in men with varicoceles, in observational studies reporting

---

**Table II. Semen analysis**

<table>
<thead>
<tr>
<th>Indices</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deterioration of: Concentration Motility Morphology Azoospermia + low α-glucosidase + normal FSH: Low fructose: seminal vesicle dysfunction High concentration of sperm surface ABS: Vitality of the single spermatozoon:</td>
<td>Subfertility (‘normality’ requires redefinition) Obstruction Immunological infertility Prognosis for ICSI</td>
</tr>
</tbody>
</table>

FSH = follicle stimulating hormone; ABS = antisperm antibodies; ICSI = intracytoplasmic sperm injection.
improvement of semen parameters after varicocelectomy, and even increased pregnancy rates. For example, a surgical approach to the varicocele problem has been justified by pointing to a 22% semen improvement rate and a 28% pregnancy rate following varicocelectomy in 82 men (Mordel et al., 1990). The same authors tabulated the results of 50 studies of the fertility effects of occlusion of the spermatic vein in men with varicoceles (Mordel et al., 1990). Of the 5471 men reported in these studies, 57% (range: 0–92%) showed improvement of semen parameters after surgery, and 36% (range: 0–63%) succeeded in achieving a pregnancy in their partners. Ideally to appraise the effect of varicocelectomy, a comparison group without treatment should be included in studies of this nature. None of the 50 studies did. So it remains difficult to establish to what extent regression towards the mean will have been responsible for the semen improvement, and whether a 36% spontaneous pregnancy rate is to be expected for this group of patients with otherwise unexplained subfertility, without surgery. Taylor and Collins (1992), in a review article of 20 studies involving 2026 couples with unexplained subfertility, calculated a 33% crude spontaneous pregnancy rate after expectant management, a figure not different from the 36% reported previously in 5471 couples (Mordel et al., 1990).

To answer the question whether varicocelectomy improves fertility in men with varicoceles, randomized controlled trials should be performed (Sackett et al., 1991). An extensive Medline search for the years 1966–1998 plus handsearching of the 26 major journals in this field revealed that only five studies of this kind have been published so far (Table III).

The combined results of these (in 417 patients) lead to the conclusion that varicocelectomy does not improve the pregnancy rate in couples with unexplained subfertility of whom the male partner has a varicocele. A systematic review of the five studies revealed that one study (Madgar et al., 1995) differed significantly from the rest. A cumulative Mantel–Haenszel meta-analysis (Table III) showed that the pooled odds ratio (OR) for pregnancy occurring after varicocelectomy as compared to no varicocelectomy was 0.93 (95% CI 0.57–1.52).

A vast body of evidence in the literature suggested a deleterious effect of varicoceles on fertility and recommending their ligation, either prophylactically or as a treatment of decreased fertility. An epiphenomenon, varicocele and sub-

<table>
<thead>
<tr>
<th>Reference</th>
<th>n</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nilsson et al. (1979)</td>
<td>96</td>
<td>0.34</td>
<td>0.07–1.42</td>
</tr>
<tr>
<td>Breznik et al. (1993)</td>
<td>175</td>
<td>0.27</td>
<td>0.12–0.61</td>
</tr>
<tr>
<td>Madgar et al. (1995)</td>
<td>220</td>
<td>0.84</td>
<td>0.42–1.67</td>
</tr>
<tr>
<td>Yamamoto et al. (1996)</td>
<td>305</td>
<td>0.81</td>
<td>0.45–1.47</td>
</tr>
<tr>
<td>Nieschlag et al. (1998)</td>
<td>430</td>
<td>0.93</td>
<td>0.57–1.52</td>
</tr>
</tbody>
</table>

n = cumulative number of patients in review; OR = odds ratio for pregnancy occurring after varicocelectomy; 95% CI = 95% confidence interval of OR.

fertility coexisting as two independent consequences of a common, yet unknown, denominator, cannot be excluded. Few randomized controlled clinical trials have been performed to study the effect of obstructing the left spermatic vein on subsequent fertility in men with varicoceles. The few, small studies are heterogeneous but they do not support a possible aetiological role of the varicocele in the existing subfertility, or at least do not support a beneficial effect of varicocelectomy. The fact that most of them show an improvement in semen analysis parameters but not in pregnancy rate suggests that the semen improvement is due to a statistical effect, regression towards the mean.

Recommendations

(i) The effect of obstructing the left internal spermatic vein (LISV) in subfertile and/or oligozoospermic men with a varicocele should be assessed through an appropriately designed, large prospective randomized study.

(ii) There is insufficient scientific evidence at the moment for recommending obstruction of the LISV routinely as treatment in subfertile and/or oligozoospermic men with a varicocele.

In summary

Many varicocele studies are of poor quality. A few are good, but small in size. They show an improvement in semen analysis parameters. They do not show an improvement in pregnancy rates. This suggests that the semen improvement may be due to ‘regression towards the mean’.

Treatments for endometriosis

Since randomized, double-blind controlled trials demonstrated the efficacy of endometriosis ablation in increasing the pregnancy rate in infertile women (Marcoux et al., 1997) and in reducing pelvic pain in symptomatic patients (Sutton et al., 1994), it may hardly be considered ethical to refrain from surgery in the presence of a laparoscopic diagnosis solely with the aim of evaluating the effects of new drugs on non-treated disease. As a result of the increasing surgical approach to endometriosis (Crosignani et al., 1996), the combination of medical treatment with laparoscopic procedures, either pre- or postoperatively, represents a growing field of application of drugs (Vercellini et al., 1998).

The quality of the evidence supporting the use of medical and/or conservative surgery for endometriosis is manifestly poor, and no recommendations can be made based on the results of the published studies (Thomas and Cooke, 1987; Telimaa, 1988; Tulandi and Mouchawar, 1991; Fedele et al., 1992; Adamson et al., 1993; Vercellini et al., 1998).

There seems to be enough available evidence to suggest that drugs suppressing ovulation are of no benefit to infertile women with endometriosis (Vandekerckhove et al., 1993; ESHRE Capri Workshop, 1996) and their use only delays potential conceptions in comparison with expectant management or alternative therapies such as assisted reproduction (Hughes et al., 1993). In the light of this scenario, hormonal drugs should no longer be prescribed, either alone or in combination with surgery, with the aim of increasing the pregnancy rate in infertile women.
**Assisted reproductive technology**

Although >100 000 IVF cycles were performed in 1993, in many countries less than one cycle of IVF per hundred infertile couples is achieved per annum. Put in other words, the uptake of IVF, which is the most effective treatment for infertility, is only a tiny fraction of the defined need. Despite this appearance of insufficient services, some couples in countries with unlimited public funding and couples with sufficient public funding in other countries will have six, eight, nine or even more cycles of IVF treatment. Is this a sensible course of action for couples and a good use of health care resources? This section will address approaches to clinical reasoning on the value of the ‘nth’ IVF cycle.

**What are the issues?**

From the patient’s viewpoint the question appears to be uncomplicated: at what point does the (constant) risk exceed the (diminishing) benefit. Of course, the question is not that simple. For example, one school of thought holds that the benefit is constant. Also, the predicted risk may decline for a woman who has undergone IVF cycles without an untoward event. Furthermore, if we take a step back from the clinical arena, pregnancy is not the only benefit that patients can accrue from IVF and the risk of complications is not the only impediment to continuing with IVF treatment.

From the viewpoint of the paying agent the question is complicated by resource constraints. Note that even when the couple uses their own money, it is lost to other opportunities. Nevertheless, for the paying agent, the question of how many IVF cycles to fund also hinges on the risk/benefit equation. Even the most beneficent society would not want to provide unlimited cycles if the procedure entailed more risk than benefit.

A further issue is the wide range of prognostic factors among couples presenting for IVF treatment. Some of these we are able to determine, such as female age. Others, such as difficulty with stimulation or fertilization, may become known during the course of one or two IVF cycles. The limited ability to predict the results of a given IVF cycle, however, suggests that many prognostic factors remain unknown. Diagnosis is not a powerful prognostic factor, unless of course there is severe male infertility, which in turn would be remedied by sperm injection, but clinical reasoning has a more secure scientific foundation when the diagnosis is one which is clearly amenable to the treatment. The specific diagnosis in this case is tubal obstruction. For other diagnoses, IVF is an empirical treatment. The value that a couple perceive for the ‘nth’ cycle may depend on whether the treatment is specific or empirical. At some future time these individual variations may be understood sufficiently to be taken into account, but for the present, understanding the prognosis for IVF means working with the average prognosis.

**The definition of risks and benefits**

Undoubtedly the dominant health benefit from IVF is the birth of a child. The non-clinical benefits also may be substantial: even if they do not conceive, the couple gains new information about infertility; they may feel that they have done everything possible to have a child, and the process may help them to reach closure. In addition to the clinical and non-clinical benefits, the process of care in the provision of IVF services also may be beneficial. When the process of care ensures autonomy in decision-making and is based on respect for individuals, undergoing IVF therapy can help to restore dignity and confidence and provide a sense of control (Ryan and Donaldson, 1996).

The downside of IVF can also be stated in terms of the health risks, non-clinical impediments and problems arising from the process of care. The notable health risks are multiple gestation pregnancy, ovarian hyperstimulation syndrome and, rarely, procedural misadventures. Non-clinical impediments include lost time and opportunities, disappointment, and continuing uncertainty. Problems arising in the process of care include long waiting times, coping with the IVF staff attitudes, and dealing with changing care-givers.

Given the number of elements on either side of the benefit/risk scales, it is not surprising that comprehensive studies are uncommon (Ryan and Donaldson, 1996). Establishing the truth of the balance is further complicated because the main health benefit, the clinical pregnancy rate, is the only element which has been studied in detail, and even these studies have been the source of debate.

**Estimating the clinical pregnancy rate in the ‘nth’ treatment cycle**

**Modelling approaches.** Predictive models have adopted one of two assumptions: that the likelihood of success remains constant, cycle after cycle; alternatively, that there would be a declining benefit. The constant model implies that the effectiveness of IVF is independent of clinical factors. The declining benefit model assumes that those who have the best prognosis conceive and are withdrawn from the pool in succeeding cycles. The predictions of most models have been unsatisfactory (Stolwijk et al., 1998).

**Empirical observations.** Once again there are contrasting views. For example, in North America a declining success rate generally pertains (Neumann et al., 1994). Data from Europe have suggested a constant success rate (FIVNAT, 1993; Croucher et al., 1998; Roest et al., 1998). A constant rate in successive IVF cycles may occur because of active censoring of patients who do poorly in a preceding IVF cycle. This withdraws both good and bad prognosis subjects from the pool in succeeding cycles. Active censoring is not a common practice in North America.

A recent study combined modelling assumptions and empirical observations. Land et al. (1997) estimated the effect of active censoring on cumulative pregnancy rates through three cycles (Land et al., 1997). They assumed that the pregnancy rate for the actively censored patients would be zero, and included these patients in the life table estimates. This approach indicated that active censoring inflates pregnancy rates by ~14%.

**The barrier to estimating success rate in the ‘nth’ cycle**

Land et al. (1997) referred to the failure to continue with IVF treatment as passive censoring. In their data 77% of couples who could have continued with IVF treatment had dropped out after the third cycle. A group of multi-centre reports involving 47 477 cycles did not segregate active
and passive censoring, and 78% of couples who started treatment and did not conceive had dropped out after the 4th cycle (Meldrum et al., 1998). Thus more than 75% of candidates for repeated IVF cycles discontinue treatment. The remainder constitute only a small fraction of the initial cohort, a condition which is fatal to the assumptions needed to construct reliable cumulative event rates. That may be no more than a statistical nicety, but clearly for the purpose of clinical or public policy it would be unwise to draw inferences about the prospects for success in repeated IVF cycles, based on the experience of this small and potentially atypical fraction of the initial cohort. Thus, although the largest body of evidence available suggests that IVF success declines in repeated cycles, an accurate estimate of the true success rate in the ‘n’th cycle of IVF treatment is not possible.

The published data also indicate that little is known of the reasons for such low continuation rates with IVF treatment. Given that drop-out rates seem to be independent of costs, it might seem that side-effects would be the main barrier to continuation of treatment. Another possibility, however, is that non-health benefits assist such couples with coming to closure on infertility treatment. The issue would be a good subject for a qualitative research study.

References
Infertility diagnosis and treatments


ESHRE Capri Workshop Group


Received on May 27, 1999; accepted on November 12, 1999