Defining poor ovarian response during IVF cycles, in women aged <40 years, and its relationship with treatment outcome

C.Kailasam1,3, S.D.Keay2, P.Wilson1, W.C.L.Ford1 and J.M.Jenkins1

1Centre for Reproductive Medicine, Division of Obstetrics and Gynaecology, University of Bristol, St Michael’s Hospital, Southwell St, Bristol, BS2 8EG and 2Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
3To whom correspondence should be addressed. E-mail: Chandra.Kailasam@repromed-bristol.co.uk

BACKGROUND: Poor ovarian response limits IVF success but assessing interventions is difficult because of the wide variation in definition. This study attempts to derive objective definitions of poor response. METHODS: A retrospective study of a consecutive series of 1190 patients aged <40 years undergoing their first IVF/ICSI cycle was undertaken. Factors adversely affecting implantation, including advanced female age, were excluded. Clinical outcome in cycles reaching oocyte retrieval (n = 1036) were evaluated with respect to gonadotrophin dose used and oocyte number. Cancelled cycles (n = 154) were analysed in relation to the stimulation dose at cancellation and outcome of their subsequent cycle. RESULTS: Cycle cancellation for patients on >300 IU FSH/day compared to those on a lower dose was associated with a significantly worse outcome in the subsequent cycle. If <3000 IU FSH/cycle were administered, clinical pregnancy rates remained favourable if <4 eggs were recovered (29 versus 33% for ≥5 eggs). By contrast, if ≥3000 IU FSH was required, the pregnancy rate was 25% if ≥5 eggs were recovered but declined to 7% if <4 were obtained. CONCLUSIONS: Definitions of poor response should include the degree of ovarian stimulation used. A low oocyte number is only detrimental if the cumulative dose is >3000 IU FSH. Cancellation at ≥300 IU FSH/day is associated with a significantly worse prognosis and could define poor response.

Key words: definition/gonadotrophins/IVF/poor ovarian response

Introduction
Poor ovarian response to gonadotrophin remains a significant problem in assisted conception. Advanced age (Akande et al., 2002), previous ovarian surgery (Nargund et al., 1995), pelvic adhesions (Keay et al., 1998a) and high body mass index (Crosignani et al., 1994; Loh et al. 2002) may be associated with poor ovarian response. However, unexpected poor responses do occur (Keay et al., 1997) and may reflect early ovarian ageing (Nikolaou and Templeton, 2003). Reliably predicting poor response may permit modification of the stimulation regime. Screening tests studied include: growth hormone (Stone and Marrs, 1992), insulin-like growth factor-I (Keay et al., 2003), Müllerian inhibitory substance (Seifer et al., 2002), inhibin B (Seifer et al., 1997), antral follicle count (Thomas et al., 1997) and ovarian volume (Lass et al., 1997), but serum FSH remains the most widely used (Toner et al., 1991; Cahill et al., 1994; Akande et al., 2002). However, intercycle variation limits both sensitivity and specificity of a single serum FSH level (Scott et al., 1990).

The number of embryos available directly influences IVF success rates (Templeton et al., 1996) and poor ovarian response reduces pregnancy rates in both the index and subsequent IVF cycles (Jenkins et al., 1991). Furthermore, simply using higher gonadotrophin doses may increase follicular recruitment but does not necessarily translate into higher pregnancy rates (Land et al., 1996). The inverse relationship between total gonadotrophin dose used and IVF success is well established (Stadtmauer et al., 1994) but the threshold at which the decline, in implantation and pregnancy rates, occurs is not. The stimulation dose used may therefore be important in considering any definition of poor response.

The lack of a uniform definition of ‘poor response’ makes it difficult to compare treatment outcomes or, indeed, develop and assess protocols for prevention and management (Surrey and Schoolcraft, 2000). The aim of the study was to define poor response with respect to measures of ovarian response (total gonadotrophin dose used, oocyte number and daily stimulation dose) and clinical outcome (embryo implantation and clinical pregnancy rates) in women who reached oocyte retrieval. Furthermore, women cancelled because of a poor response were studied in their next stimulation cycle and the outcome related to the stimulation dose at cancellation in the index cycle.
Materials and methods

Patients and protocols
A retrospective study reviewing the clinical records of a consecutive series of 1190 patients undergoing their first cycle of IVF or ICSI treatment during a 4 year period from 1997 to 2000 at the Centre for Reproductive Medicine, University of Bristol was undertaken. Inclusion in the study was limited to women aged <40 years, with two ovaries, no uterine anomaly or ultrasonically visible hydrosalpinx. The study was conducted within the guidelines of the local research ethics committee and women gave written informed consent for all clinical procedures.

Stimulation protocol
The stimulation protocol and laboratory methods have been described in detail previously (Keay et al., 1998b). In brief, norethisterone 5 mg twice daily was administered orally for 7 days from the 19th day of the preceding cycle to reduce the incidence of functional ovarian cysts (Aston et al., 1995). Pituitary desensitization was started in the mid-luteal phase of the cycle preceding IVF treatment, using intranasal buserelin acetate spray (Suprefact; Hoesscht, UK) 600 µg daily in five divided doses, 100 µg at 4-hourly intervals during the day and 200 µg at bedtime. Serum estradiol was measured after 2 weeks of buserelin administration. An estradiol level of <200 pmol/l (corresponding to the 5th centile in our population) was considered to confirm ovarian suppression.

If the estradiol level was <200 pmol/l, exogenous FSH was administered. The dose of FSH was 150 IU in patients aged <35 years and 300 IU in patients 35 years (Menogon, Ferring, UK; Gonal-F, Serono, UK). The choice of gonadotrophins, i.e. Gonal-F or Menogon, was taken by the patient and was entirely voluntary. Transvaginal ultrasonography was used to monitor follicular growth, starting on stimulation day 8 and repeated as necessary. hCG (5000 IU) was administered and transvaginal oocyte retrieval undertaken if ≥3 follicles of 18 mm were recruited. The decision to cancel treatment was discussed with the patient if <3 follicles of size 18 mm were recruited. If only 1 or 2 follicles had been recruited in response to gonadotrophins, it was considered more appropriate to cancel the cycle and start again on a higher dose rather than increase the dose during treatment. Oocyte retrieval was performed under ultrasound guidance and two or three embryos were transferred as per patient choice. Luteal phase support was achieved with cyclogest pessaries.

Mature oocytes were defined by appearance according to our laboratory guidelines and constituted oocytes typically with well-expanded cumulus, which is stretchy and silvery in appearance and a partially or fully expanded corona radiata with the outline of the oocyte visible. Oocytes with little or non-uniform expansion of the cumulus and a dense aggregation of cells in the corona were regarded as immature and excluded from the analyses.

Patients were asked to perform a pregnancy test 2 weeks after embryo transfer if they had not menstruated. The cycle was deemed to be successful only if the fetal heart was demonstrable on ultrasound scan. Cycles cancelled due to a poor response and therefore not reaching oocyte retrieval were analysed with respect to the dose of gonadotrophin used. Patients whose cycles had been cancelled on <300 IU for poor follicle recruitment had their FSH dose increased to 300 IU in the subsequent cycle. In general, patients cancelled on 300 IU FSH did not have a dose increase in their subsequent stimulation cycle although some women opted to have 450 IU FSH, despite the lack of evidence to support this dose increase.

The clinical outcome of their next ovarian stimulation cycle was recorded and analysed in relation to the dose of gonadotrophin used in the cancelled cycle.

Fertilization was defined as normal by the development of two pronuclei and progressive cleavage up to the time of embryo transfer after 2–3 days. Implantation rates were defined by the proportion of individual embryos transferred resulting in a gestation sac, including ectopic gestations with and without a fetal heartbeat. Clinical pregnancy was indicated by an intrauterine gestation sac in which a fetal heartbeat could be seen on ultrasound examination 4 weeks after embryo transfer. The number of sacs present and pregnancy outcome were also recorded.

Measures of ovarian response
In 1036 cycles reaching oocyte retrieval (recruitment of three pre-ovulatory follicles ≥18 mm) the total gonadotrophin dose per cycle, the number of oocytes retrieved and the gonadotrophin dose per oocyte were analysed in relation to individual embryo implantation and clinical pregnancy rate.

In cycles cancelled because of inadequate follicular development (n = 154), the clinical outcome in their next stimulation attempt was related to the daily gonadotrophin stimulation dose at the time of cancellation.

Assays
Serum estradiol (E2) was assayed using a radioimmunoassay (Delfia, Wallac, UK). The intra-assay and inter-assay imprecision (coefficients of variation) of the E2 assay were 7.1 and 9.7% respectively.

Statistical analysis
The Mann–Whitney U-test, Student’s t-test, χ²-test and Fisher’s exact test were used for comparison between groups where appropriate using the statistical package StatsDirect (StatsDirect Ltd, UK). Multi-way χ²-tables were analysed using the G-test (Sokal and Rolhf, 1969). P < 0.05 was considered statistically significant.

Results
Women cancelled on <300 IU FSH achieved comparable pregnancy rates, in their next stimulation attempt, to women who reached oocyte retrieval in their initial stimulation cycle, during the same period (22.0 versus 28.1% respectively, P = 0.07). Cancellation at a dose of ≥300 IU FSH was associated with a significantly worse outcome than cancellation at a lower daily dose (clinical pregnancy rate 6.7 versus 22.0%, P < 0.05 and repeat cancellation for poor response 24.4 versus 6.4% respectively, P < 0.005) (Table I).

Patients were banded into response groups by the total FSH requirement and a significant decline in oocyte number with increasing cumulative FSH dose was observed (Table II).

In patients where the ovarian response was considered to be adequate, leading to oocyte retrieval, pregnancy rates were favourable when <3000 IU FSH (32.4% clinical pregnancy rate) were administered irrespective of the number of oocytes collected (Table III). If >3000 IU FSH were required, there was a significant reduction in overall clinical pregnancy rate [21.9 (≥3000 IU FSH) versus 32.4% (<3000 IU FSH), P < 0.0005]. It is evident (Table III) that if ≥3000 IU FSH is required, the oocyte number becomes critical. When ≥3000 IU FSH was used, pregnancy rates were very poor if <4 oocytes were collected but declined even if >4 oocytes were collected (6.5 versus 25.4% respectively, P < 0.0005). Both clinical pregnancy and implantation rates declined significantly with increasing FSH requirement per oocyte recovered (Figure 1).

Discussion
This is a pragmatic study assessing ovarian response in a standard IVF programme using a long GnRH agonist protocol.
Table I. Outcome of subsequent IVF cycle following cancellation because of inadequate follicular development

<table>
<thead>
<tr>
<th>FSH (IU)</th>
<th>Daily dose at cancellation</th>
<th>Repeat cancellation rate</th>
<th>Clinical pregnancy rate</th>
<th>Live-birth rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;300 IU FSH</td>
<td>7/109 (6.4)</td>
<td>24/109 (22.0)</td>
<td>22/109 (20.2)</td>
<td></td>
</tr>
<tr>
<td>≥300 IU FSH</td>
<td>11/45 (24.4)</td>
<td>3/45 (6.7)</td>
<td>2/45 (4.4)</td>
<td></td>
</tr>
<tr>
<td>P (Fisher’s exact test)</td>
<td>&lt; 0.005</td>
<td>&lt; 0.05</td>
<td>&lt; 0.05</td>
<td></td>
</tr>
</tbody>
</table>

Values in parentheses are percentages.

Table II. Numbers of mature oocytes recovered from women <40 years during their first cycle of IVF according to the cumulative dose of FSH

<table>
<thead>
<tr>
<th>FSH (IU)</th>
<th>No. of women</th>
<th>Median</th>
<th>Quartiles</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1500</td>
<td>56</td>
<td>11</td>
<td>7–16</td>
<td>3–29</td>
</tr>
<tr>
<td>1575–2250</td>
<td>418</td>
<td>11</td>
<td>7–15</td>
<td>1–36</td>
</tr>
<tr>
<td>2325–3000</td>
<td>148</td>
<td>10</td>
<td>7–14</td>
<td>1–37</td>
</tr>
<tr>
<td>3075–3750</td>
<td>203</td>
<td>9</td>
<td>6–14</td>
<td>1–48</td>
</tr>
<tr>
<td>3825–4500</td>
<td>170</td>
<td>7</td>
<td>5–11</td>
<td>1–29</td>
</tr>
<tr>
<td>≥4500</td>
<td>41</td>
<td>7</td>
<td>5–10</td>
<td>1–20</td>
</tr>
</tbody>
</table>

Kruskal–Wallis test ($\chi^2$) = 58.048; $P < 0.0001$.

in women aged <40 years. The potential to achieve pregnancy following cancellation for an inadequate response (<3 developing pre-ovulatory follicles) in the first treatment cycle was dependent on the daily FSH dose at cancellation. Women cancelled, on <300 IU FSH/day, realized pregnancy rates similar to normal responders, reflecting a failure to reach the threshold for multiple follicle recruitment. However, cancellation on ≥300 IU FSH/day was associated with a significantly higher repeat cancellation rate and a reduced pregnancy rate (24.4 versus 6.4%, and 6.7 versus 22.0% respectively). These findings suggest that inadequate follicular development in response to 300 IU FSH daily identifies women whose reproductive performance is limited by the capacity of the ovary to respond. Conversely an initial poor response to <300 IU FSH is unlikely to be detrimental, and simply increasing the stimulation dose to 300 IU FSH/day will result in a satisfactory ovarian response in most cases.

Previously we have cautioned against the use of daily stimulation doses >300 IU FSH as clinical pregnancy rates are very low (<5%) despite more pre-ovulatory follicles being recruited (Karande et al., 1990; Jenkins et al., 1991; Manzi et al., 1994; Land et al., 1996). The live-birth rate is likely to be significantly lower given the association between diminished ovarian reserve and increased early pregnancy loss (Levi et al., 2001) and using >300 IU FSH daily appears difficult to justify as the costs incurred are directly proportional to the amount of FSH used.

Analysis was restricted to the first cycle of treatment, which eliminated bias that may arise from previous treatment or from the inclusion of multiple cycles, and women with factors known to adversely affect embryo implantation were excluded (advanced age, uterine anomaly and hydrosalpinx). We did not attempt to determine outcome in relation to screening tests (such as basal FSH) but used the actual response, as gonadotrophin stimulation prior to IVF has been demonstrated to be a sensitive, dynamic test of ovarian reserve (Beckers et al., 2002). Moreover, a poor response may precede the development of abnormal screening tests (Farhi et al., 1997).

Several strategies have been investigated to improve treatment outcome but have met with limited success (Karande and Gleicher, 1999; Surrey and Schoolcraft, 2000; Keay, 2002). We have reported the benefit of adjuvant low dose dexamethasone in reducing the incidence of poor response prior to IVF, and incorporating this into standard stimulation regimes may improve the efficiency of gonadotrophin stimulation (Keay et al., 2001) but its efficacy in critically defined poor responders has not been tested.

It is clear that women requiring high amounts of gonadotrophin to produce an average oocyte yield have a significantly reduced chance of conceiving. The gonadotrophin requirement, a direct indicator of the ovarian reserve, appears to identify qualitative differences in the oocytes produced. There is a continuous relationship between the gonadotrophin required per oocyte retrieved and clinical pregnancy rate rather than a discrete threshold where treatment outcome declines. Furthermore, women requiring <3000 IU FSH to reach oocyte retrieval had satisfactory pregnancy rates even when <4 oocytes were retrieved. This is consistent with the observation that young women with normal ovarian reserve who recruited few follicles achieved pregnancy rates that were sure to adversely affect embryo implantation were excluded the inclusion of multiple cycles, and women with factors known to adversely affect embryo implantation were excluded (advanced age, uterine anomaly and hydrosalpinx). We did not attempt to determine outcome in relation to screening tests (such as basal FSH) but used the actual response, as gonadotrophin stimulation prior to IVF has been demonstrated to be a sensitive, dynamic test of ovarian reserve (Beckers et al., 2002). Moreover, a poor response may precede the development of abnormal screening tests (Farhi et al., 1997).

Several strategies have been investigated to improve treatment outcome but have met with limited success (Karande and Gleicher, 1999; Surrey and Schoolcraft, 2000; Keay, 2002). We have reported the benefit of adjuvant low dose dexamethasone in reducing the incidence of poor response prior to IVF, and incorporating this into standard stimulation regimes may improve the efficiency of gonadotrophin stimulation (Keay et al., 2001) but its efficacy in critically defined poor responders has not been tested.

It is clear that women requiring high amounts of gonadotrophin to produce an average oocyte yield have a significantly reduced chance of conceiving. The gonadotrophin requirement, a direct indicator of the ovarian reserve, appears to identify qualitative differences in the oocytes produced. There is a continuous relationship between the gonadotrophin required per oocyte retrieved and clinical pregnancy rate rather than a discrete threshold where treatment outcome declines. Furthermore, women requiring <3000 IU FSH to reach oocyte retrieval had satisfactory pregnancy rates even when <4 oocytes were retrieved. This is consistent with the observation that young women with normal ovarian reserve who recruited few follicles achieved pregnancy rates that were
broadly similar to women with a higher response (Lashen et al., 1999). We suggest that a total dose of $\geq 3000$ IU FSH in this study population (particularly when four or fewer oocytes are collected) objectively identifies a poor response in women aged $<40$ years. Furthermore, using oocyte number to define poor response, without reference to stimulation dose used, may be misleading as a low oocyte yield may not prejudice treatment if $<3000$ IU FSH is used. This is important as the definition of poor response varies (Keay et al., 1997) and oocyte number, without reference to gonadotrophin dose used, has been widely used to identify poor responses. Couples should be made aware that an encouraging number of oocytes collected and fertilized does not necessarily lead to an improved outcome.

In summary, we suggest that in women aged $<40$ years, poor response should be defined as cancellation on $\geq 300$ IU FSH daily ($<3$ developing pre-ovulatory follicles) or a requirement to administer $\geq 3000$ IU FSH in total to recruit sufficient follicles to justify oocyte retrieval. Even then it should be recognized that there is a continuum of ovarian responsiveness that limits the usefulness of these definitions. Research on poor ovarian response has been hampered by a lack of definitions; thus we hope this work may contribute to the development of an agreed definition(s) for poor ovarian response feeding into a current initiative by ESHRE to develop a European Classification for Infertility (http://www.ECT.info).

References
Keay SD, Liversedge NH, Akande VA, Mathur RS and Jenkins JM (2003) Serum insulin-like growth factor 1 concentrations following pituitary desensitisation do not predict the ovarian response to gonadotrophin stimulation prior to IVF. Hum Reprod 18,1797–1801.

Submitted on January 23, 2004; accepted on March 25, 2004