

Three-dimensional ultrasound improves the interobserver reliability of antral follicle counts and facilitates increased clinical work flow

K. JAYAPRAKASAN, B. K. CAMPBELL, J. S. CLEWES, I. R. JOHNSON and N. J. RAINE-FENNING

Nottingham University Research and Treatment Unit in Reproduction (NURTURE), Academic Division of Reproductive Medicine and Surgery, School of Human Development, Queen's Medical Centre, University of Nottingham, Nottingham, UK

KEYWORDS: antral follicle count; interobserver reliability; ovarian reserve; three-dimensional ultrasound; transvaginal ultrasound; work flow

ABSTRACT

Objectives To compare the interobserver reliability of antral follicle counts made using real-time two-dimensional (2D) ultrasound with offline counts made from stored three-dimensional (3D) data and to assess the time required for such counts.

Methods Two observers conducted transvaginal ultrasound examinations in 45 subfertile women in the early follicular phase of the menstrual cycle. Antral follicles were counted using real-time 2D ultrasound and the time taken was recorded. A 3D volume was then acquired from each ovary and stored for subsequent offline analysis using the multiplanar view. The time taken for each step was recorded and the total time was calculated. Intraclass correlation coefficients (ICC) and limits of agreement were used to assess reliability.

Results There was no difference in the mean antral follicle counts using real-time 2D (16.51 ± 11.51) and 3D (16.33 ± 12.13) ultrasound. According to ICCs, there was a significantly higher interobserver reliability for counts made using 3D (mean, 0.992; 95% CI, 0.986–0.996) compared with real-time 2D (mean, 0.961; 95% CI, 0.940–0.977) ($P < 0.01$) ultrasound. 3D ultrasound was also associated with narrower limits of agreement (-2.7 to $+3.1$) than was 2D ultrasound (-6.9 to $+6.4$). Whilst the total time taken was significantly longer for the 3D technique (239.3 ± 71.4 s vs. 103.1 ± 28.6 s, $P < 0.001$), the time required for the actual ultrasound examination was significantly less (46.4 ± 7.4 s vs. 103.1 ± 28.6 s, $P < 0.001$).

Conclusions 3D ultrasound significantly improves the interobserver reliability of antral follicle counts. While this is at the expense of time overall, the duration of the actual ultrasound examination and patient exposure is significantly reduced using 3D compared with real-time 2D ultrasound. Copyright © 2008 ISUOG. Published by John Wiley & Sons, Ltd.

INTRODUCTION

The possibility of conception, either spontaneously or in conjunction with fertility treatment, declines steadily with age¹. This age-related decrease in a woman's fertility potential is due primarily to a progressive decline in the ovarian reserve², which is defined as the number and quality of primordial follicles that remain within the ovaries³. There is considerable variation in ovarian reserve between different women of the same age. This reflects differences in the number of primordial follicles present at the start of reproductive life and in the timing of the age-related decline in their number, which is determined genetically but influenced by many environmental factors⁴.

Accurate, reliable assessment of ovarian reserve is essential in women undergoing assisted reproduction treatment as it allows prediction of their probable response to ovarian stimulation and potential modification of their treatment to maximize this. Whilst tests of the ovarian reserve are often normal and reassuring, they occasionally reveal unexpected results. A small, but significant, number of young women have diminished ovarian reserve and thus a reduced chance of successful outcome, whilst

Correspondence to: Dr K. Jayaprakasan, Nottingham University Research and Treatment Unit in Reproduction (NURTURE), Academic Division of Reproductive Medicine and Surgery, School of Human Development, Queen's Medical Centre, University of Nottingham, Nottingham, Nottinghamshire NG7 2UH, UK (e-mail: k.jayaprakasan@nottingham.ac.uk)

Accepted: 12 February 2008

an even smaller proportion of older women have an unexpectedly good ovarian reserve and a relatively good chance of conception⁵. Accurate estimation of the ovarian reserve facilitates pre-treatment counseling of couples and assists the clinician to formulate an individualized treatment plan.

Several endocrine and ultrasound markers of ovarian reserve have been proposed and adopted into clinical practice over the last 15 years. All of these tests primarily aim to estimate the number of gonadotropin-responsive or 'selectable' follicles, which are assumed to be reflective of the primordial follicle population⁶. Assessment of ovarian reserve is made directly using ultrasound (antral follicle count⁷ and ovarian volume⁸) or indirectly through serum measurements of the endocrine factors produced by the developing follicles (estradiol⁹, inhibin B¹⁰ and anti-Mullerian hormone (AMH)¹¹) or the hormones under the inhibitory control of these factors (follicle stimulating hormone (FSH)¹², luteinizing hormone¹³). All of these tests have different sensitivities, specificities and predictive values, and there is a lack of consensus as to the best single test or combination of tests⁵. The number of antral follicles measured using ultrasound appears to be superior to most endocrine tests^{14,15} and ovarian volume measurements¹⁶ in predicting the response to ovarian stimulation during assisted reproduction treatment cycles. Serum AMH levels appear equally predictive¹¹, but their routine use is limited by the lack of a reliable cut-off level and by cost. There is a great deal of debate as to the best test of ovarian reserve, but in the absence of multicenter, prospective, randomized controlled trials, the antral follicle count is considered the test of choice for quantifying ovarian reserve¹⁶.

Counting of the antral follicles can be performed with conventional, real-time, two-dimensional (2D) ultrasound⁷ or with three-dimensional (3D) ultrasound. In contrast to 2D, 3D ultrasound provides the user with a variety of different image displays, which facilitates a range of different measurement techniques^{17,18}. However, 3D ultrasound requires the acquisition of volumetric data and the subsequent offline analysis of these data, which may increase the overall time required for assessment. A single study has compared the reproducibility of antral follicle counts made with real-time 2D and 3D ultrasound¹⁹. This study was limited by its design, which allowed the recruitment of two different study groups and involved the use of two different ultrasound machines, one for each method, either of which could have affected the image quality. A similar study by our group compared the reliability of three counting techniques, including a '2D-equivalent method' using stored 3D data¹⁷, but a direct comparison between real-time 2D and 3D techniques was not performed.

This study is the first in which a comparable subject group has been examined, using the same ultrasound machine, to assess the interobserver reliability of antral follicle counts using conventional, real-time 2D ultrasound and stored 3D volume datasets, acquired independently by two different observers. This study is

also the first to consider the effect of the number of antral follicles on count reliability and to report the time required for data acquisition and counting in the context of the clinical setting, in order to examine the effect that 3D ultrasound examination might have on work flow.

METHODS

Experimental design

We recruited 51 consecutive subjects undergoing investigation for subfertility who were under 40 years of age, with regular menstrual cycles of 21 to 35 days' duration and an early follicular phase FSH level of < 10 IU/L and who had both ovaries. All subjects underwent a baseline ultrasound assessment in the early follicular phase (day 2–5) of the menstrual cycle. Subjects were excluded if they had a history of previous ovarian surgery or were found to have an ovarian cyst or follicle measuring 20 mm or more in diameter. The study was approved by the hospital's local ethics committee and informed written consent was obtained prior to the enrolment of each subject.

Data acquisition and counting

All subjects had a transvaginal scan performed independently by each of two observers (K.J. and J.C.) in a sequential manner using a Voluson Expert 730TM (General Electric Medical Systems, Zipf, Austria) ultrasound machine equipped with a four-dimensional 7.5-MHz transvaginal probe. A probe program (Gain, 5; SRI, 3; Enhance, 1; Reject, 15; Harmonics, high) that subjectively provided the best 2D gray-scale image was loaded for each patient and the settings maintained throughout the study. Each subject was asked to empty their bladder and was then scanned with their legs supported by stirrups in a modified Lloyd Davies position, to limit discomfort and to ensure free manipulation of the transvaginal transducer.

A routine 2D ultrasound assessment of the pelvis was first performed by one of the two observers to exclude any obvious pathology. The transducer was positioned to show the longitudinal view of the uterus for the 2D ultrasound assessment of the ovaries and for acquisition of the 3D volume datasets in order to set a reference point from which to measure the time taken for both techniques. With the 2D technique the ovaries were visualized in the longitudinal plane and the number of antral follicles measuring 2–10 mm in diameter within each ovary was counted as the observer slowly moved the transducer from one side of the ovary to the other. The time taken to complete a single count of the total number of antral follicles in both ovaries, from the reference starting point, was noted. Two more counts were performed by the same observer so that three values were obtained for each ovary. The mean of these three values is presented in the results section but the time shown relates to that needed for a single count.

A 3D acquisition of each ovarian volume was then obtained using the slow sweep mode, as described

previously²⁰. The volume of interest and speed of acquisition were maintained throughout the study and were identical for every subject. The resultant multiplanar display was examined to ensure that the ovary had been captured in its entirety. A single acquisition was obtained for each ovary and the data were saved to the hard drive of the ultrasound machine. The total time taken from the reference starting point to the completion of data acquisition and storage was recorded.

The second observer, who was unaware of the other observer's results, then repeated the transvaginal scan and performed the antral follicle count using the 2D technique before acquiring 3D volumes of both ovaries, as described above. The total time taken for each step was recorded in the same manner. The order that each observer examined the subject was determined randomly. A single observer was present during each ultrasound assessment. The data were saved using nomenclature that did not disclose the person who had performed the ultrasound examination, but coded to allow precise identification of the order and observer if required.

Volumetric data were subsequently transferred to a personal computer via a DVD without data compression. The stored 3D data were analyzed at least 2 weeks after the initial real-time assessment to avoid the influence of recall effects. The counts were performed on a personal computer using 4D View (version 6.0; General Electric Medical Systems). Independently, the two observers (K.J. and J.C.) counted the total number of antral follicles three times in each ovarian volume dataset using all three perpendicular planes that are shown simultaneously in the multiplanar view (Figure 1). The order of counting was determined randomly and was different for each observer. Both ovaries from the same subject were analyzed together and the total time taken by each observer to arrive at a final total antral follicle count for each subject was recorded.

Statistical analysis

Statistical analysis was performed using the Statistical Package for the Social Sciences (SPSS version 14.0,

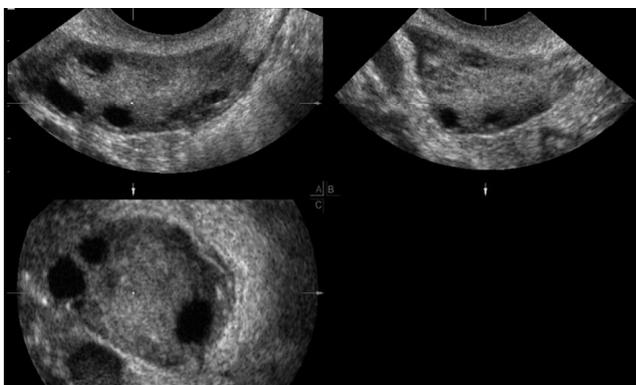


Figure 1 Multiplanar display of a three-dimensional ovarian volume dataset demonstrating the three mutually related orthogonal image planes: longitudinal view or A-plane (upper left); transverse view or B-plane (upper right) and coronal view or C-plane (lower left).

Chicago, IL, USA). The total antral follicle count obtained by adding the number of follicles in each ovary was considered for analysis. Interobserver reliability for both 2D and 3D techniques was assessed by two-way mixed intraclass correlation coefficients (ICCs) with absolute agreement and their 95% CIs²¹. The difference between pairs of ICCs was assessed using Fisher's *z* transformation, with significance determined using the *t*-statistic. The mean antral follicle count was calculated from the three repeated counts of each subject for each observer. These pairs of mean counts were used to calculate the mean differences and the limits of agreement (LOA) between the observers for each method and these values are presented as Bland–Altman plots²². The paired *t*-test was used to examine for significant differences in the mean antral follicle counts between the observers and between each counting method for each observer. A paired *t*-test was also used to examine the difference in the time taken for the count between each method. *P* < 0.05 was considered statistically significant.

RESULTS

A total of 51 subjects were recruited and all of them underwent a pre-treatment transvaginal ultrasound examination. Six of these subjects were excluded because they had ovarian follicles or cysts measuring more than 20 mm in diameter. The final study group of 45 subjects was included in the final analysis. The mean (\pm SD) age of these subjects was 34.4 (\pm 3.6) years and their mean basal FSH level was 7.21 (\pm 2.04) IU/L. The median number of antral follicles was 12 (range, 2–53). There was no difference in the mean antral follicle count between the two different ultrasound techniques for either observer, either individually or in combination (Table 1).

The mean ICCs for both ultrasound techniques were all above 0.961, both within and between observers, which is indicative of a high degree of reliability. However, antral follicle counts made from stored 3D data had a significantly greater (*P* < 0.01) ICC and narrower 95% CI than had those made using the 2D technique (Table 2).

Bland–Altman plots comparing the reliability of the two counting techniques are shown in Figure 2. Whilst the mean difference in the antral follicle count between observers for the 2D technique (–0.29) and the 3D technique (0.19) were similar and small, the range

Table 1 Comparison of antral follicle counts on real-time two-dimensional ultrasound (2D) and using stored three-dimensional ultrasound data (3D)

Technique	Antral follicle count (mean \pm SD)		
	Observer 1	Observer 2	Observers 1 & 2
2D	16.36 \pm 11.21	16.67 \pm 12.02	16.51 \pm 11.51
3D	16.47 \pm 12.42	16.19 \pm 11.97	16.33 \pm 12.13
<i>P</i>	NS	NS	NS

NS, not significant.

Table 2 Interobserver reliability (assessed by intraclass correlation coefficient) of antral follicle counts on real-time two-dimensional ultrasound (2D) and using stored three-dimensional ultrasound data (3D)

Technique	Intraclass correlation coefficient (mean (95% CI))		
	Observer 1	Observer 2	Observers 1 & 2
2D	0.981 (0.961–0.990)	0.983 (0.966–0.991)	0.961 (0.940–0.977)
3D	0.989 (0.979–0.994)	0.993 (0.989–0.996)	0.992 (0.986–0.996)
P	NS	NS	< 0.01

NS, not significant.

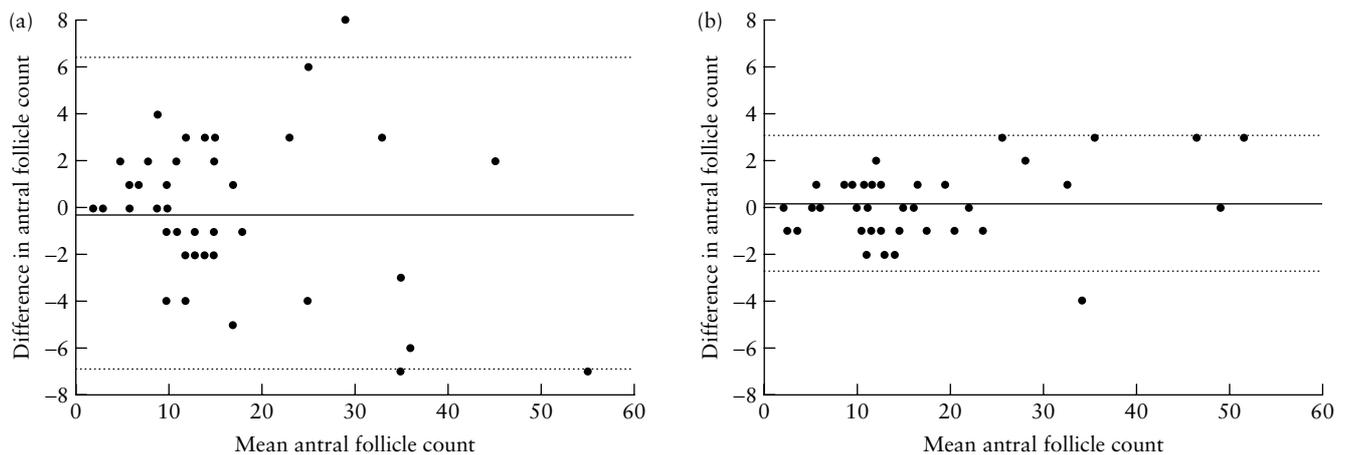


Figure 2 Interobserver reliability of antral follicle counts: Bland–Altman plots showing difference between two measurements against their mean for counts made on real-time two-dimensional ultrasound (a) and from stored three-dimensional ultrasound data (b). Lines indicate mean difference (—) and limits of agreement (.....).

of differences between the observers differed for each method and were narrower for the 3D technique. The 3D technique was associated with LOA of -2.7 to $+3.1$, indicative of a high degree of reliability between observers, whilst the comparable figures for the 2D technique were -6.9 to $+6.4$.

Subgroup analysis was performed to assess the effect of the total number of antral follicles on count reliability. Subjects were categorized into four subgroups according to each quartile of the total antral follicle count. There was a distinct trend towards a decrease in reliability with both ultrasound techniques as the number of follicles increased, as indicated by the increasing range between the upper and lower LOA (Table 3). However, the 3D technique showed much better reliability across all subgroups compared with real-time 2D ultrasound.

The mean (\pm SD) time required to perform a single count of the total number of antral follicles within both ovaries of one subject using real-time 2D ultrasound was 103.1 ± 28.6 s per subject (Table 4) and the mean (\pm SD) time taken to acquire 3D volumes of both ovaries was 46.4 ± 7.4 s per subject. Therefore, as expected, each subject spent significantly ($P < 0.001$) less time in the ultrasound room for the 3D technique than they did for the real-time 2D evaluation of the ovaries. However, subsequent off-line analysis of the 3D data required a further 192.9 ± 67.6 s per subject to perform the total antral follicle count. In combination, the total time

Table 3 Interobserver reliability (assessed by limits of agreement (LOA)) of antral follicle counts made with real-time two-dimensional (2D) and three-dimensional (3D) ultrasound techniques according to the absolute number of follicles

Antral follicle count	Mean difference (LOA)	
	2D	3D
2–10	0.91 (–1.79; 3.6)	0.18 (–1.29; 1.65)
11–13	–0.42 (–6.23; 5.4)	0.19 (–2.21; 2.59)
14–21	–1.46 (–6.38; 3.47)	–0.38 (–2.83; 2.08)
> 21	–0.53 (–11.59; 10.52)	1.1 (–3.41; 5.61)

required to acquire and measure volumetric data from both ovaries was 239.3 ± 71.4 (mean \pm SD) s, which was significantly ($P < 0.001$) longer than the time required for the real-time 2D technique.

DISCUSSION

This is the first study to compare the interobserver reliability of antral follicle counts made with real-time 2D and 3D ultrasound techniques and to assess the time required for such counts. While antral follicle counts using both 2D and 3D ultrasound were highly reliable, assessment of the total antral follicle count using 3D ultrasound data was significantly more

Table 4 Time required to perform antral follicle counts on real-time two-dimensional ultrasound (2D) and three-dimensional ultrasound (3D)

Parameter	Time (s)		
	Mean	SD	Range
2D ultrasound			
Time	103.1	28.6	57–179
3D ultrasound			
Time for acquisition	46.4	7.4	35–70
Time for count	192.9	67.6	86–330
Total time	239.3	71.4	121–400

reliable, as indicated by a significantly higher ICC and narrower LOA.

Our findings are contradictory to those reported in the only other study comparing the reproducibility of follicle counts using real-time 2D and 3D ultrasound. In the study by Scheffer *et al.*¹⁹, the reproducibility of antral follicle counts was examined with real-time 2D in 37 volunteers with proven fertility and from 49 3D ultrasound datasets stored from prior ultrasound assessments of general fertility patients. Whilst the interobserver reproducibility was found to be adequate (ICC for both 2D and 3D, 0.98) with both techniques, their 3D technique did not reveal any advantage over the real-time 2D method in terms of count reliability (LOA for 2D, -5.0 to $+4.1$ and for 3D, -5.6 to $+5.7$). Our study design ensured that the same population was examined with both techniques, in a prospective manner. While one might assume the two groups studied by Scheffer *et al.* were comparable, the study groups were too small to account for population differences such as varying body mass index and pelvic anatomical relationships or to account for the inherent differences in image quality seen between any two patients. Furthermore, while we used the same ultrasound machine and identical settings for both 2D and 3D assessments in order to reduce differences in image quality between the two techniques, Scheffer *et al.* used two different machines, which introduced another source of bias into their study.

Both observers in our study acquired their own 3D volume datasets for each ovary, resulting in two datasets per subject. These were analyzed independently by both observers, blinded to the subject and the acquirer. This strengthens the findings of our study, in terms of a true comparison of the reliability between the two techniques, as both acquisition and data counting were considered. Inclusion of more than two observers would have added more power to our study, but this was limited by the practical difficulty of scanning each subject in real-time at the same sitting by multiple observers. A compromise of two was considered adequate and allowed assessment of interobserver reliability. Both observers in our study were experienced in 2D and 3D transvaginal ultrasound and had considerable experience in the assessment of antral follicle counts. Our findings of improved reliability of 3D over 2D ultrasound are therefore likely to reflect a

valid test result that carries significant implications with regard to the accurate assessment of ovarian reserve in the clinical setting. The relatively short learning curve in post-processing of stored 3D data for follicle counts means the 3D technique offers a high degree of count reproducibility even with relatively inexperienced observers¹⁷.

A high degree of reliability is required for the assessment of antral follicle counts in all patients undergoing assisted reproduction treatment, but it is absolutely essential in those with a limited ovarian reserve because it predicts probable poor responders to ovarian stimulation. Identification of patients with a low total antral follicle count facilitates protocol modification to ensure a maximal response in the first cycle and allows appropriate counseling of the couple prior to treatment. Recent work from our Assisted Conception Unit has demonstrated that women with an antral follicle count of seven or less, as measured using either a 3D or a 2D-equivalent method from stored 3D data, are at significantly increased risk of poor response (sensitivity, 100%; specificity, 92.6–93.6%), defined as either cycle cancellation or retrieval of fewer than four oocytes²³. As expected, in our current study, count reliability was better in subjects with lower numbers of antral follicles than it was in those with high follicle counts with either counting technique. However, antral follicle counts made with real-time 2D ultrasound demonstrated significantly wider LOA even at lower total follicle counts, suggesting that 3D ultrasound is a better tool for assessment of these patients. This could have important clinical implications with regard to the pre-treatment counseling of women and their individual treatment, but this needs to be tested in randomized, prospective studies. The precision of follicle counts in patients with a high number of antral follicles may not be as clinically important, but it will allow for a reliable objective diagnosis of polycystic ovaries (PCO) and those women likely to have an exaggerated response to ovarian stimulation. Whilst the long-term implications of isolated PCO in the absence of oligo/amenorrhea or hyperandrogenism are uncertain²⁴, the presence of PCO constitutes a significant risk factor for ovarian hyperstimulation syndrome during *in-vitro* fertilization treatment²⁵. In our study population, according to both observers with 2D and 3D ultrasound, there were eight (17.8%) subjects with bilateral PCO and one (2.2%) with unilateral PCO, diagnosed on the criteria defined at the joint consensus meeting of the American Society of Reproductive Medicine and the European Society of Human Reproduction and Embryology²⁶. Although the reproducibility at higher antral follicle counts was significantly worse with 2D compared with 3D ultrasound, it is reassuring that none of the PCO cases was missed by either of the observers or the techniques.

The time taken for antral follicle counts could be an important factor in determining their reliability. Compared with 2D ultrasound, the 3D technique took significantly longer to arrive at a final antral follicle count, but subjects were in the examination room for

less than half the time required for the real-time 2D assessment. With real-time 2D ultrasound, counting must be performed instantaneously from a single image plane, in the presence of the patient. In contrast, 3D ultrasound facilitates the storage of volume data which can be analyzed subsequently, in the absence of the patient, in a virtual real-time manner using a simultaneous display of three orthogonal image planes. Technicians could be trained to do this, allowing a more appropriate and focused use of resources, with important implications for work flow in busy units. While the infrastructure would need to be modified to account for the additional time required for subsequent analysis, this should be balanced against the fact that 3D data can be stored for a virtual real-time assessment of the patient at any stage in the future and could form the basis of an educational library. The use of 3D ultrasound facilitates an improvement in clinical efficiency and patient throughput, with enhanced reliability of antral follicle counts.

3D ultrasound appears to confer clinical benefits over 2D ultrasound for the assessment of antral follicles, although its routine use in clinical practice is currently limited by cost. However, as has happened with Doppler, with time 3D imaging will probably become integrated into most ultrasound machines as a standard. The 3D technique has considerable potential in research, educational and clinical settings. Its implications in fertility patients and for the work flow within assisted conception units needs to be tested in randomized, prospective trials.

REFERENCES

- Baird DT, Collins J, Egozcue J, Evers LH, Gianaroli L, Leridon H, Sunde A, Templeton A, Van Steirteghem A, Cohen J, Crosignani PG, Devroey P, Diedrich K, Fauser BC, Fraser L, Glasier A, Liebaers I, Mautone G, Penney G, Tarlatzis B; ESHRE Capri Workshop Group. Fertility and ageing. *Hum Reprod Update* 2005; **11**: 261–276.
- Faddy MJ, Gosden RG, Gougeon A, Richardson SJ, Nelson JF. Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. *Hum Reprod* 1992; **7**: 1342–1346.
- Maheshwari A, Fowler P, Bhattacharya S. Assessment of ovarian reserve—should we perform tests of ovarian reserve routinely? *Hum Reprod* 2006; **21**: 2729–2735.
- te Velde ER, Pearson PL. The variability of female reproductive ageing. *Hum Reprod Update* 2002; **8**: 141–154.
- Broekmans FJ, Kwee J, Hendriks DJ, Mol BW, Lambalk CB. A systematic review of tests predicting ovarian reserve and IVF outcome. *Hum Reprod Update* 2006; **12**: 685–718.
- Gougeon A. Regulation of ovarian follicular development in primates: facts and hypotheses. *Endocr Rev* 1996; **17**: 121–155.
- Tomas C, Nuojuua-Huttunen S, Martikainen H. Pretreatment transvaginal ultrasound examination predicts ovarian responsiveness to gonadotrophins in in-vitro fertilization. *Hum Reprod* 1997; **12**: 220–223.
- Lass A, Skull J, McVeigh E, Margara R, Winston RM. Measurement of ovarian volume by transvaginal sonography before ovulation induction with human menopausal gonadotrophin for in-vitro fertilization can predict poor response. *Hum Reprod* 1997; **12**: 294–297.
- Smotrich DB, Widra EA, Gindoff PR, Levy MJ, Hall JL, Stillman RJ. Prognostic value of day 3 estradiol on in vitro fertilization outcome. *Fertil Steril* 1995; **64**: 1136–1140.
- Seifer DB, Scott RT, Jr., Bergh PA, Abrogast LK, Friedman CI, Mack CK, Danforth DR. Women with declining ovarian reserve may demonstrate a decrease in day 3 serum inhibin B before a rise in day 3 follicle-stimulating hormone. *Fertil Steril* 1999; **72**: 63–65.
- van Rooij IA, Broekmans FJ, te Velde ER, Fauser BC, Bancsi LF, de Jong FH, Themmen AP. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. *Hum Reprod* 2002; **17**: 3065–3071.
- Sharif K, Elgendy M, Lashen H, Afnan M. Age and basal follicle stimulating hormone as predictors of in vitro fertilisation outcome. *Br J Obstet Gynaecol* 1998; **105**: 107–112.
- Weghofer A, Feichtinger W. The forgotten variable: impact of luteinizing hormone on the prediction of ovarian reserve. *Fertil Steril* 2006; **85**: 259–261.
- Bancsi LF, Broekmans FJ, Eijkemans MJ, de Jong FH, Habbema JD, te Velde ER. Predictors of poor ovarian response in in vitro fertilization: a prospective study comparing basal markers of ovarian reserve. *Fertil Steril* 2002; **77**: 328–336.
- Hendriks DJ, Mol BW, Bancsi LF, Te Velde ER, Broekmans FJ. Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level. *Fertil Steril* 2005; **83**: 291–301.
- Hendriks DJ, Kwee J, Mol BW, te Velde ER, Broekmans FJ. Ultrasonography as a tool for the prediction of outcome in IVF patients: a comparative meta-analysis of ovarian volume and antral follicle count. *Fertil Steril* 2007; **87**: 764–775.
- Jayaprakasan K, Walker KF, Clewes JS, Johnson IR, Raine-Fenning NJ. The interobserver reliability of off-line antral follicle counts made from stored three-dimensional ultrasound data: a comparative study of different measurement techniques. *Ultrasound Obstet Gynecol* 2007; **29**: 335–341.
- Raine-Fenning NJ, Lam PM. Assessment of ovarian reserve using the inversion mode. *Ultrasound Obstet Gynecol* 2006; **27**: 104–106.
- Scheffer GJ, Broekmans FJ, Bancsi LF, Habbema JD, Looman CW, Te Velde ER. Quantitative transvaginal two- and three-dimensional sonography of the ovaries: reproducibility of antral follicle counts. *Ultrasound Obstet Gynecol* 2002; **20**: 270–275.
- Raine-Fenning NJ, Campbell BK, Clewes JS, Johnson IR. The interobserver reliability of ovarian volume measurement is improved with three-dimensional ultrasound, but dependent upon technique. *Ultrasound Med Biol* 2003; **29**: 1685–1690.
- Khan KS, Chien PF. Evaluation of a clinical test. I: assessment of reliability. *BJOG* 2001; **108**: 562–567.
- Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995; **346**: 1085–1087.
- Jayaprakasan K, Hilwah N, Kendall NR, Hopkisson JF, Campbell BK, Johnson IR, Raine-Fenning NJ. Does 3D ultrasound offer any advantage in the pretreatment assessment of ovarian reserve and prediction of outcome after assisted reproduction treatment? *Hum Reprod* 2007; **22**: 1932–1941.
- Hart R, Hickey M, Franks S. Definitions, prevalence and symptoms of polycystic ovaries and polycystic ovary syndrome. *Best Pract Res Clin Obstet Gynaecol* 2004; **18**: 671–683.
- Engmann L, Maconochie N, Sladkevicius P, Bekir J, Campbell S, Tan SL. The outcome of in-vitro fertilization treatment in women with sonographic evidence of polycystic ovarian morphology. *Hum Reprod* 1999; **14**: 167–171.
- Balen AH, Laven JS, Tan SL, Dewailly D. Ultrasound assessment of the polycystic ovary: international consensus definitions. *Hum Reprod Update* 2003; **9**: 505–514.