Variation in the determination of follicular diameter: an inter-unit pilot study using an ultrasonic phantom

G.L. Driscoll1, J.P.P. Tyler1 and D. Carpenter2
1 CityWest IVF, 12 Caroline St, Westmead, NSW 2145, Australia
2CSIRO Division of Telecommunication and Industrial Physics, Lindfield NSW 2070, Australia

To whom correspondence should be addressed

Ultrasound operators in assisted reproductive technology units in New South Wales, Australia cooperated in a study that was conducted to assess their ability to measure fixed objects embedded in an ultrasound phantom. The results have demonstrated a large variation ranging between 10 and 25% coefficient of variation on distances between 10 and 32 mm. As ultrasonic imaging is the only component of assisted reproductive technology not currently controlled by an external quality assurance scheme, the authors suggest there may be a need to establish a programme.

Key words: assisted reproductive technology/follicular diameter/quality control/ultrasound

Introduction

The monitoring of ovarian follicular development by transvaginal ultrasound is a standard procedure in assisted reproductive technology treatment. Furthermore with the increasing use of ‘down-regulation’ with gonadotrophin releasing hormone (GnRH) analogues (e.g. lucrin, buserelin, syneral) the endocrine component of follicular tracking has been reduced and clinical decisions have become more dependent on the results of the ultrasound scans.

The optimum follicle size from which to induce ovulation is not precisely known and the decision to give human chorionic gonadotrophin (HCG) depends on a number of factors including the total number of follicles recruited, their echogenic appearance by ultrasound (Fukuda et al., 1995; Gore et al., 1995), the sizes of the leading group and the perceived potential for hyperstimulation. Thus different clinics have their own criteria. This ranges between two or more follicles of >17 mm diameter to those >21 mm (personal communication with all Australian assisted reproductive technology units). Considerable debate continues as to which is the ‘correct’ approach and about whether one is demonstrably better than another. However, according to statistics compiled by the Fertility Society of Australia (Lancaster et al., 1997) most Australian units appear to be achieving similar results with respect to pregnancy success, perhaps suggesting a wide window for inducing ovulation pre-assisted reproductive technology. Another reason for this wide latitude may be the calibration and/or operator variation in the ultrasound assessment of follicles.

There have been few studies examining the performance of ultrasonic follicular measurements in assisted reproductive technology and those that have been done have examined observer reproducibility using a single scanner (Forman et al., 1991) rather than a range of machines likely to be found in different units. Studies in other disciplines (e.g. obstetrics) have, however, demonstrated the need for experienced ultrasonographers (Bahmaie and Edmonds, 1996) and the necessity of instrument calibration (Clark, 1988; Fish, 1990; Dudley and Griffith, 1996). Similarly there is increasing discussion about quality assurance in diagnostic medicine (Arger, 1995; Henderson, 1996; Weber, 1996) as the number of ultrasound services per 1000 head of population in Australia has increased from 34.9 to 97.9 over a 4 year period (Hailey, 1996). Similarly statements from professional bodies allied to ultrasound in the UK suggest that training and performance review should be routine and a Consortium for the Accreditation of Sonographic Education (CASE) has been established to provide an overview of the education of sonographers (BMUS Bulletin, 1996).

Thus the object of this preliminary project was to assess the variation in ultrasonic determination of objects approximating follicular sizes using a standard ultrasound phantom distributed between seven assisted reproductive technology units and to make recommendations on the need — or otherwise — of establishing an external quality assurance (EQA) programme for ultrasound operators in assisted reproductive technology units.

Materials and methods

Experimental design

For this preliminary study a standard ‘phantom’ used for the calibration of ultrasonic instruments (ATS Multipurpose Phantom Model 539: Gamasonics Institute for Research and Calibration, PO Box 411, Mittagong, NSW 2575, Australia) was used. Measurements were not made of the pre-set patterns within the phantom which would have alerted the operator to the probable ‘correct’ value required.

The ultrasound phantom

The phantom is a solid, ultrasound-permeable block of rubberized material in which are embedded wires and cylinders which, when scanned by ultrasound, reflect the appearance of either white spots or black filled circles of different diameters and consistency against a background echo scattering designed to represent liver tissue. They are situated at different levels and therefore can be used as both far and near objects (relative to the scanning probe). The technique of scanning a phantom is quite critical and the transducer needs to be aligned correctly to give the clearest picture. Similarly maximizing adjustments to the machine (e.g. total gain settings) will give clearer imaging. These, of course, are factors which would also affect the scanning and visualization of human ovaries.
Participants
The phantom was taken to each unit in Sydney and one in Canberra (see Acknowledgements for a list of the participating assisted reproductive technology units) and every operator who performed ultrasound measurements for that programme given the opportunity to take part in the study. The type of instrument used, the probe frequency and the status of the operator (trained ultrasonographer, medical practitioner, nurse specialist etc.) was also recorded. Each was shown the layout of the phantom before measurements were made.

Measurements
Ten values were requested (reproduced in Figure 1 and indexed as measurements 1–10 in the results). These were distances between different-sized ‘circles’, so had no exact distance by which a ‘wanted’ value could be guessed. Within the 10 determinations three sets of paired measurements were made and the final value (measurement 10) was a calibrated distance of 2 cm. Thus an assessment of an instrument’s calibration and inter-observer variation in assessing distance could be made. Each operator performed the measurements without knowledge of the values obtained by the others.

Figure 1. A diagram (not to scale) of the layout of the phantom. Participants were requested to measure the distances defined by the numbered ‘lines’ (i.e. outer circumference to outer circumference for measurements 3, 6 and 8; inner circumference to inner circumference for measurements 2, 5 and 9; diameter for measurement 4, etc.).

Results
Twenty-four operators from seven units completed the measurements. Five were trained ultrasonographers, 10 were experienced clinical nurse specialists, seven were experienced medical practitioners and two were scientists with no experience in ultrasound methods. The instruments used were: Aisonics Opus 1 with 7.5 MHz probe (10 operators), Aisonics Micro with 7.5 MHz probe (three operators), Siemens with 5.0 MHz probe (two operators), Acuson 128XP4 with 5.0 MHz probe (one operator), RT 3600 with 5.0 MHz probe (one operator), GERT with 5.0 MHz probe (one operator), Toshiba 140 with 7.5 MHz probe (one operator), ATL50 with 5.0 MHz probe (one operator) and Sonoace 1500 with 7.5 MHz probe (four operators).

A summary of the results for each requested measurement is given in Table 1. Mean values were normally distributed (mean approximated median value) and the measurements (mean range 10–32 mm) approximated the clinical situation in determining follicular diameter. However, the coefficient of variation was large and ranged between 10.1 and 29.9%. Furthermore it appeared unrelated to the size of the measurement (correlation of linear regression, \( r = 0.38 \)) or to the distance of the object from the probe.

For those measurements that were replicated (3 vs 8; 3 vs 6; 6 vs 8; 2 vs 5) no statistically significant difference could be found \( (P > 0.05) \) using Student’s paired \( t \)-test. While the numbers were not large it was also possible to suggest that probe type (5.0 or 7.5 MHz) and operator experience also did not markedly affect measurement performance. However some experienced nurse specialists did have problems in clearly defining the ‘deeper’ and smaller measurements (numbers 1, 2, 6 and especially 5) suggesting a lack of knowledge of the ability to optimize their machine settings. Of interest was that measurements generated by the two novice participants always fell within the interquartile range.

The distribution of values for each measurement is diagrammatically represented as box plots in Figure 2 where ‘outliers’ can be identified and the notched box represents the interquartile range (25th to 75th percentiles). Thus even with a defined distance (measurement 10 which was factory set at 20 mm) the range was between 18.0 and 29.4 mm. A line plot

<table>
<thead>
<tr>
<th>Table 1. Summary statistics (in centimetres) for each measurement requested in the pilot external quality assurance study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement no.(^a)</td>
</tr>
<tr>
<td>No. of operators(^a)</td>
</tr>
<tr>
<td>Mean distance</td>
</tr>
<tr>
<td>SD</td>
</tr>
<tr>
<td>Coefficient of variation (%)</td>
</tr>
<tr>
<td>Median value</td>
</tr>
<tr>
<td>Minimum value</td>
</tr>
<tr>
<td>Interquartile range</td>
</tr>
<tr>
<td>Maximum value</td>
</tr>
</tbody>
</table>

\( \text{aRefer to Figure 1.} \)

\( \text{bWhere } n \text{ does not equal } 24 \text{ some operators were unable to assess a measurement because of technical difficulties.} \)
Ultrasound variation in assisted reproductive techniques

Figure 2. Notched box plots of all data for each measurement (refer to Figure 1) showing median values (notch), the interquartile range (box), 10th and 90th percentiles (whiskers) and outliers.

of the data (Figure 3) further suggests that no one particular observer or instrument consistently affected the variation seen in these data.

Discussion

All aspects of pathology laboratories in Australia are required to be accredited by the National Agency of Testing Authorities (NATA). An integral component of this registration is demonstration of competence in an external quality assurance scheme. Similarly the Australian assisted reproductive technology licensing body [the Reproductive Technology Accreditation Committee (RTAC)] also requires careful monitoring of in-vitro fertilization culture laboratories, treatment consent forms, patient satisfaction with information brochures and counselling practice etc. Ultrasound measurement of ovarian follicles remains the only ‘uncontrolled’ aspect of assisted reproductive technology treatment.

Notwithstanding the question of suitability of tissue-mimicking phantoms in assessing equipment calibration (Letter to the Editor, 1985), and the importance of the observer as a significant source of variability in clinical data assessment (Corson et al., 1995), this pilot survey has demonstrated a large variation in the ability of the participants to determine distance in an ultrasound phantom. It could be argued that this is of little consequence for assisted reproductive technology units since pregnancy rates across Australia appear similar (Lancaster et al., 1995) and an ‘eyeball’ approach to follicular size and number may be all that is necessary in clinical practice, particularly to limit the potential for ovarian hyperstimulation. However, the corollary of this is that the phantom is made using stable rigid bodies and thus should not be as difficult to measure as the ovary where it might be argued that variation could be even larger.

Thus an error of 18% would mean that a follicle actually measuring 18 mm could be anywhere between 14.7 and 21.2 mm. While the exact relationship between follicle volume and oocyte competence is still not precisely known, it is widely accepted that an oocyte collected from a smaller follicle (<16 mm) would generally be less competent than that from a larger one (>18 mm). Similarly, experienced ultrasonographers are capable of distinguishing healthy from atretic follicles but admit that it is ‘impractical to observe and assess more than four or five follicles in one ovary by ultrasound’, a scenario often encountered in women stimulated for assisted reproductive technology (Fukuda et al., 1995).

Furthermore, intracytoplasmic sperm injection (ICSI) is demonstrating that there is as much variation in oocyte quality as there is in sperm morphology and motility. Should a woman decide not to use ovulating drugs for her assisted reproductive technology attempt then the current approach (i.e. as long as there is a reasonable group of follicles ‘some’ will be competent) ceases to be clinically applicable and the need to maximize oocyte quality based on biochemical and follicular maturity becomes paramount.

Finally, in this survey the lowest coefficient of variation was 10%, a value generally considered the upper limit of allowable error in many pathology tests. The resolution of ultrasound determination is obviously limited by many factors (machine settings, patient compliance etc.) and since ovarian follicles are not usually spherical the interpretation of their volume will remain subjective. In conclusion this pilot study suggests the need for an EQA programme for ultrasound in assisted reproductive technology programmes. Individual clinics could consider making their own ‘perishable’ phantoms (Gibson and Gibson, 1995) for in-house assessment or develop within their geographic areas an external programme such as has been established for mammography (Royal Australian College of Radiology, 1996).

Acknowledgements

The authors thank Dr A.Ekangaki, of Macquarie University, for statistical assistance, the Ultrasounds Laboratory, CSIRO, for the loan of the phantom for this study and the following Medical Directors for allowing their staff to participate in this survey: Dr Geoffrey Driscoll, CityWest IVF, 12 Caroline St, Westmead. NSW 2145; Dr Robert Jansen, Sydney IVF, 4 O’Connell St, Sydney, NSW 2000; Dr Trevor Johnson, Royal Hospital for Women Fertility Group.
References


Received on February 4, 1997; accepted on July 8, 1997