The effect of follicle volume measurement on clinical decisions

Mark J Gooding\textsuperscript{a}, Debbie Barber\textsuperscript{b}, Stephen Kennedy\textsuperscript{b} and J Alison Noble\textsuperscript{a}

\textsuperscript{a}Medical Vision Laboratory, University of Oxford, UK
\textsuperscript{b}Nuffield Dept. of Obstetrics and Gynaecology, University of Oxford, John Radcliffe Hospital, UK

1 Introduction

Currently, the growth of ovarian follicles under stimulation for fertility treatment is monitored using 2D ultrasound. The follicle size is recorded as the mean of two approximately orthogonal diameters across the follicle, in the plane in which the follicle appears largest. Human Chorionic Gondaotropin (hCG) is administered to the patient 36 hours prior to oocyte recovery, at a time when the oocytes are considered to be mature. The decision as to when to administer hCG is based solely on the mean diameter measurement, subject to blood estradiol levels being within acceptable limits.

In [1] we presented a semi-automatic method for the measurement of follicle volumes from freehand 3D ultrasound. This paper presents a study of the accuracy of the volume measurements, together with a study of the effect that using this volume measurement would have on management of the hCG decision.

2 Method

2.1 Accuracy of 3D measurement

In [1] we presented a semi-automatic method for the measurement of follicle volumes from freehand 3D ultrasound. This method uses a level set based region growing algorithm to segment manually identified follicles in 3D, whilst simultaneously interpolating the surface where data is absent. To assess the accuracy of this 3D volume measurement, volumes measurements from the reconstruction system were compared to aspirated follicle volumes as follows; The ovaries from 9 patients undergoing IVF treatment were scanned prior to oocyte recovery. The entire ovary was scanned for the 3D scan and all follicles were reconstructed. To aid identification of each follicle at the time of aspiration the clinician who would perform the follicle aspiration observed the 3D scan and labelled diagrams were drawn for each scan by the scanning clinician. Still images of the significant follicles were also printed and labelled. Although ideally all follicles would be measured in each ovary, if an ovary contained more than about 6 follicles, identification of follicles at aspiration is very difficult. Therefore only the follicles for which the aspirating clinician could be confident of a correct identification had volumes measured and recorded. Aspirated volumes were recorded to the nearest 0.5ml. The current clinical measure of follicle size recorded is the mean diameter. This assumes that the follicle is spherical and therefore that the diameter is representative of the follicle volume. To test this assumption, mean diameter measurements were also made at the time of the 3D scan. The volumes of the follicles were estimated from the clinical diameter measurements using a spherical model, for comparison to the aspirated volume.

2.2 Effect of volume on clinical decision

For a new system to be adopted for use in a clinical context it is necessary to show that quality of treatment management is at least equivalent to that of the current system in use. For fertility treatment management the quality of treatment management must be assessed for the decision as to when to administer hCG.

After a period of about 1 week of monitored hormonal stimulation of the ovaries, the follicles are nearing maturity. At this stage the decision as to when to aspirate the follicles must be made. The hormone hCG must be administered 36 hours before aspiration, to complete the maturation. Currently the decision of when to administer hCG and aspirate the follicles is made largely on an assessment of the number of mature follicles. A follicle is considered mature if it is greater than 18mm in mean diameter, and is considered post-mature if greater than 25mm. The hCG decision must be generally consistent between systems, if the reconstruction system is to be used for treatment management. It must be shown that the same decision is made whether based on volume measurement or based on mean diameter. In cases of a difference of decision, it is necessary that the volume measurement provides a better tool for treatment management.

\*gooding@robots.ox.ac.uk
Figure 1. This figure shows the volume estimated from 3D reconstruction plotted against the clinician's diameter measurement. The board relationship established aids the clinician in understanding the new volume measurement in terms of the familiar diameter measurement. 2.14ml is equivalent to 18mm and 4.77ml is equivalent to 25mm.

To investigate whether the use of volume instead of diameter affects this decision, a simulation of this decision was used. Both diameter and volume measurements from 16 patients were collected. These measurements were presented to a clinician independently and in a random order. The clinician was asked to make a decision as to whether the follicles were mature and if the patient was ready for the hCG injection. The clinician was given access to the general relationship between volume and diameter to assist them in understanding the volume measurements. The measurements from all 16 patients were included randomly twice, to evaluate if the clinician’s decision was consistent. The 32 diameter and 32 volume sets were coded so that the clinician was not aware of the correspondence between diameter data sets and volume data sets, or between repeated measurements. However since the clinician was not accustomed to making an hCG decision based on volume, it was necessary to establish a parametric relationship between mean diameter and volume. This relationship was found as described in [2], by measuring the volume and diameter of 191 follicles and fitting a curve through the data as the model of the relationship. The clinician was allowed to refer to this relationship, as presented in figure 1, when presented with the volume data, such that an equivalent diameter could be found.

3 Results

3.1 Accuracy of 3D measurement

The data acquired consisted of 21 follicles for which aspirated volume, reconstructed volume, and diameter measurements were available. The low number of follicles in this part of the study is a result of the difficulty in identifying follicles at aspiration once the first follicle has been aspirated. This problem was also identified in [3], where only the largest follicle was measured in each of 25 patients.

Use of the Bland-Altman plot [4], shown in figure 2, illustrates that the estimated volume underestimates the aspirated volume by about 25%. The reasons for the underestimate have not been identified, and it is noted that the clinical measurement also underestimates the aspirated volume by about 25%. However, our main concern is with the reproducibility compared to the clinical measurement, since a measurement with a proportional error can still be used as an effective measure, but a measurement with high variance cannot.

Qualitatively it can be seen that the measurement made by the 3D system has a lower spread than the clinical data. Quantitatively, the variance of each data set from a linear fit to the data was 1.13ml$^2$ and 0.43ml$^2$ for the clinical and automated data sets respectively. For an F-test to be used to determine if the difference in variance between these sets is significant, it is necessary that the sets follow a Normal distribution. A chi-squared test was performed to check for normality. None of the data sets were rejected from being considered Normal. The F-test shows that the 3D reconstruction system has a significantly different variance to both the clinical measurements (p=0.966).
Figure 2. This graph shows the Bland-Altman plot for the reconstruction volume measurements and clinical volume estimates. Both the reconstruction method and clinical measurement have similar bias. The reconstruction method has a lower variance than the clinical measurement.

This variance is much lower than that of the clinical measurement, and therefore the 3D reconstruction system can be considered to be out-performing the current clinical measurement method.

3.2 Effect of volume on clinical decision

Generally the hCG decision made by the clinician was consistent for both presentations of each set of patient data, whether based on diameter and volume. For both diameters and volumes measurements there was only one patient (6.25%) for whom the decision of the clinician was different on the second occasion. However, this discrepancy occurred with a different patient for each measurement method. This is not unexpected since borderline cases will exist independent of the measurement method used. This experiment does show that the decision made by the clinician is highly repeatable.

For two patients (12.5%) the decision of the clinician differed when using volume measurements rather than diameter measurements. For these patients it is necessary to examine the reason for the discrepancy. In both cases the volume measurements were larger than would be expected given the diameter measurements, leading the clinician to decide on the basis of volume that hCG should be administered, whereas on the basis of diameter hCG would be postponed. Figure 3 shows the 3D reconstruction for an ovary from one of these patients, together with the plane in which the diameter was measured manually. The foreground follicle corresponds to the follicle for which the diameter measurements are shown. In the reconstruction is can be seen that the plane of measurement was perpendicular to the major axis of the follicle and as such the diameter measurement can be considered to be in error. In a protocol where the aim is to find the plane of maximum diameter it is expected that any error in finding this plane will lead to underestimation of the follicle size, leading to the hCG injection being incorrectly postponed.

4 Conclusion

It has been shown that the automated volumetric follicle reconstruction method presented in [1] has significantly lower variance than the state-of-the-art 2D clinical measurement. It should be expected that the diameter measurement would have higher variation because the diameter measurement cannot account for the shape of the follicles as they deviate from the spherical model. Despite this variation, there existed a large degree of agreement between hCG decision based on diameter measurement and those based on volume measurement. In those cases where there was a difference in the decision made, it is apparent that the volume measurement is of assistance, because the true size of the follicle is known. Therefore use of this measure may lead to better management of the treatment. However there still exists borderline decision cases even when using volume measurement. This should be expected because the decision is still fairly subjectively based on the size of the follicle and the expected oocyte yield. Use of expert systems [5] may be of benefit in such a well specified decision environment. Although it was
Figure 3. This figure shows the plane in which the clinician made the diameter measurement, together with the 3D reconstruction of the follicles within the ovary. This plane was perpendicular to the major axis of the follicle show with measurements. The equivalent follicle can be seen in the foreground on the reconstruction. In such cases the diameter measurement can be seen to be in error.

possible for the clinician to base decisions on volume measurements, there was a need to have the relationship with diameter available to help the clinician understand how to interpret the volume measurement. This need will persist to some degree when/if volume measurements are adopted clinically.

Further work should consider the effect on treatment outcome of management based on follicular volume. Variability of decision between clinicians should also be considered.

Acknowledgements

Mark Gooding is funded by the EPRSC as part of the MIAS-IRC. (GR/N14248).

References