Abstract: Our study examined AFC variation across the cycle and its impact on clinical management. We documented AFC in early (iAFC) and late follicular phase (sAFC) in 79 women. We examined absolute agreement between iAFC and sAFC as well as agreement for categorisation into categories of risk of extremes of ovarian response. We compared controlled ovarian stimulation (COS) protocols designed with iAFC and sAFC and the predictive value of iAFC and sAFC for extremes of ovarian response in women who underwent COS. We found significant difference between iAFC and sAFC (16 [IQR 9 - 23] vs. 13 [IQR 7 - 21], p=0.001) with moderate agreement for the classification into at risk of extremes of response (k=0.525). There was good agreement for protocol selection based on either AFC (k=0.750) and starting gonadotrophin dose (CCC 0.970 [95% CI 0.951 - 0.982]). iAFC and sAFC maintained good predictive value for poor ovarian response and risk of ovarian hyperstimulation syndrome (AUC 0.932 [0.832 - 1.0], 0.927 [0.850 - 1.0]) and (AUC 0.701 [0.533 - 0.869], 0.737 [0.504 - 0.971]) respectively. AFC varies across the menstrual cycle but this does not have a significant impact on COS protocol design and prediction of extremes of ovarian response.
Dear Professor Garcia-Velasco,

Please find attached our manuscript on the variation between antral follicle counts determined in the early and late follicular phases of the menstrual cycle. We aimed to describe the variation and explore whether this variation significantly impacts clinical decisions.

We look forward to hearing from you in due course.

Kind regards

Dr Dimitrios Mavrelos
Variation in antral follicle counts (AFC) at different times in the menstrual cycle – does it matter?

Short title: Variation in AFC during the menstrual cycle

Dimitrios Mavrelos Ali Al Chami, Vikram Talaulikar, Elizabeth Burt, Lisa Webber, Ephia Yasmin
Reproductive Medicine Unit, University College London Hospital, 25 Grafton Way, London NW1 2BU, United Kingdom

Corresponding author
Dimitrios Mavrelos
Reproductive Medicine Unit
University College London Hospital
25 Grafton Way
NW1 2BU
United Kingdom
d.mavrelos@ucl.ac.uk
Abstract

Our study examined AFC variation across the cycle and its impact on clinical management. We documented AFC in early (iAFC) and late follicular phase (sAFC) in 79 women. We examined absolute agreement between iAFC and sAFC as well as agreement for categorisation into categories of risk of extremes of ovarian response. We compared controlled ovarian stimulation (COS) protocols designed with iAFC and sAFC and the predictive value of iAFC and sAFC for extremes of ovarian response in women who underwent COS. We found significant difference between iAFC and sAFC (16 [IQR 9 – 23] vs. 13 [IQR 7 – 21], p=0.001) with moderate agreement for the classification into at risk of extremes of response (k=0.525). There was good agreement for protocol selection based on either AFC (k=0.750) and starting gonadotrophin dose (CCC 0.970 [95% CI 0.951 – 0.982]). iAFC and sAFC maintained good predictive value for poor ovarian response and risk of ovarian hyperstimulation syndrome (AUC 0.932 [0.832 – 1.0], 0.927 [0.850 – 1.0]) and (AUC 0.701 [0.533 – 0.869], 0.737 [0.504 – 0.971]) respectively. AFC varies across the menstrual cycle but this does not have a significant impact on COS protocol design and prediction of extremes of ovarian response.

Keywords: ovarian reserve testing, AFC, AMH, ultrasound
Introduction

Assessment of the biomarkers of ovarian response is integral to the work up of women presenting with subfertility. Biomarkers used in routine clinical include anti-mullerian hormone (AMH), early follicular phase follicle stimulating hormone (FSH) and early follicular phase antral follicle count (AFC). AMH is the marker of ovarian reserve with least intracycle variability (Deb et al. 2013) making it the most convenient as it can be performed at any point in the cycle (Iliodromiti et al. 2014). FSH is a late marker of poor ovarian response (Broekmans et al. 2006; Iliodromiti et al. 2014) and alone is not considered adequate to design a protocol of controlled ovarian stimulation (COS) (Broer et al. 2013). AFC by transvaginal ultrasound in the early follicular phase of the cycle continues to be part of initial investigations for women presenting with subfertility as the test can be easily performed directly by the treating clinician with immediately available results (Scheffer et al. 2002).

An inherent concern with use of AFC as a biomarker is its operator dependence (Iliodromiti and Nelson 2015). Another concern has been timing of AFC determination with regard to menstrual cycle. Current recommendations limit AFC determination to the early follicular phase in an attempt at standardization (Broekmans et al. 2010). However this restriction creates anxiety for patients as it creates a narrow time window within which to complete their assessments and results in considerable administrative burden for clinics to schedule appointments. It also increases the number of visits for the patient to the clinic.

The aim ovarian response biomarker determination is to identify patients at risk extremes of ovarian response and individualize COS protocols to obtain an optimal result (Bosch and Ezcurra 2011). A degree of intracycle variability that does not jeopardize these objective may acceptable in clinical practice. Indeed some data on this subject exist, a retrospective study of over 3,000 patients argued that the clinical usefulness of AFC remains unchanged across the cycle (Rombauts et al. 2011).
The aim of our study was to examine whether AFC determination in the late follicular phase of the cycle would impact selection of COS protocol and the accuracy of AFC to predict extremes of ovarian response.

Materials and Methods

The study took place in the Reproductive Medicine Unit (RMU) of University College London Hospital (UCLH) between April 2014 and June 2015. We sought the opinion of the Joint Research Office of the hospital and were advised that formal ethics approval was not required as the project involved no change in routine clinical practice.

We included women with a regular 28 – 34 day cycle referred for fertility investigation. All women referred to the clinic undergo a transvaginal ultrasound scan between days 2 to 5 of their menstrual cycle for examination of the pelvis, uterus and assessment of the AFC (Voluson E8 Expert, GE Medical Systems, Zipf, Austria). All women undergo a second ultrasound examination on day 8 – 12 of their cycle for Hysterosalpingo – Contrast - Sonography (HyCoSy) or 3D saline infusion sonohysterography (3D SIS). At the second examination the ovaries are routinely examined and a repeat AFC is performed. Determination of the antral follicle count is performed according to internationally agreed guidance (Broekmans et al. 2010). Briefly, we identify each ovary and examine it in two planes to determine its limits. We then sweep the complete ovary in the transverse plane to identify and count all follicles 2 – 10mm in diameter. Abdominal pressure is applied in cases of difficult visualization. All examinations were performed by three operators (DM, AA, and VT).

Further routine investigations in our unit for women with subfertility are serum AMH (Beckman Coulter AMH Gen II ELISA) and day 2 -5 FSH determinations. AMH and FSH results were archived on a database which is not accessed during ultrasound examinations thereby blinding operators to these results.

We created a spreadsheet (Excel 11 for Mac, Microsoft Corp.) to record the women’s demographic details, day of the cycle the first examination took place
and the early follicular phase AFC (iAFC) in each ovary. We created a second spreadsheet to document the day of the cycle the second examination took place and the late follicular phase AFC (sAFC) in each ovary as well as the presence and mean diameter of a dominant follicle in either ovary. Once all examinations were complete the databases were amalgamated and the AMH and FSH results were added.

We used the National Institute of Clinical Excellence (NICE) fertility guideline to define categories of “at risk of low response”, “normal response” and “at risk of high response” after COS (NICE 2013) i.e.: i) total AFC ≤4 and ≥16 ii) AMH ≤5.4pmol/l and ≥25.0 respectively. We classified women into the various categories based on each of AMH, iAFC and sAFC.

During the study we excluded women with amenorrhea or irregular cycles as we would not be able to time AFC determination, women with known pathologies such as endometriosis or large fibroids that displace the ovaries and affect the accuracy of AF counts. We also excluded women over 40 years old, those with a BMI>30 and those who did not tolerate HyCoSy or 3D SIS examination.

In order to examine the interobserver variability in AFC determination in our unit, a set of ovarian 3D volumes were obtained by DM and AA from a different cohort of patients and stored. The volumes were anonymised and each operator was asked examine the volumes and record the total AFC for each woman in individualized Excel spreadsheets. Each operator was blind to other operators’ findings.

COS protocols in our unit are individualized according to age, body mass index (BMI), AMH, AFC, FSH and clinician preference (Bosch and Ezcurra 2011). To further explore the potential impact of performing AFC in the late follicular phase criteria we created two Excel spreadsheets with clinical and ORT data for each case. Both spreadsheets contained age, AMH, FSH and a total AFC value. One spreadsheet contained the iAFC and the other sAFC. The cases were arranged in random order and there were no identifiers. EY was blinded as to whether the
spreadsheet contained iAFC or sAFC and was asked to select a COS protocol. The protocols we routinely use include: (i) long agonist, (ii) long agonist with withdrawal of GnRHa on day 3 of stimulation, (iii) antagonist and (iv) no COS. EY also selected starting dose of gonadotrophin dose (human menopausal gonadotrophin, hMG) for each case.

For women who underwent a COS cycle we collected data on IVF protocol, total gonadotrophin dose, serum oestradiol concentration at trigger and the total number of oocytes retrieved. We classified women who had ≤4 oocytes retrieved as “poor response” and ≥16 oocytes retrieved as “at risk of ovarian hyperstimulation syndrome (OHSS)”.

Statistical analysis

Interobserver variability was assessed by calculating Lin’s concordance correlation coefficient (CCC), plotting Bland – Altman plots and calculating the absolute and relative limits of agreement (LoA) between observers. To assess agreement between iAFC and sAFC we calculated Lin’s CCC, plotted Bland – Altman charts and calculated LoA. To examine the impact of AFC timing on i) classification into categories of predicted response to COS and ii) on COS protocol selection we constructed 2xn tables to calculate Cohen’s κ for category agreement between iAFC and sAFC. To further assess the impact of AFC timing on COS protocol selection we calculated the CCC for gonadotrophin starting dose in COS protocols designed based on iAFC and sAFC. To assess the diagnostic value of iAFC, sAFC and AMH for the identification of poor responders to COS and those at risk of OHSS we constructed receiver operating curves (ROC) and calculated the area under the curve (AUC). We used SPSS for statistical analysis. (SPSS statistics, ver. 22, IBM. Corp.)

Results

Inter-observer variability

During the study period bilateral 3D ultrasound ovarian volumes were collected from 24 women. There was good or very good interobserver agreement for off-
line AFC determination. The CCC for total AFC determination between DM and AA was 0.967 (95% CI 0.926 – 0.985) and 0.992 (95% CI 0.983 – 0.996) between DM and VT. The LoA for DM and AA was between -6.7 and 8.6 or -14%/6% of the mean AFC. The LoA for DM and VT was between -1.5 and 4.0 or -5%/14% of the mean AFC (Figure 1 and 2).

iAFC and sAFC agreement
During the study period 103 women underwent both early and late follicular phase total AFC determination by one of the observers. 24 women were excluded because the interval between examinations was >6 months. The median age of women participating in the study was 35 years (IQR 32 – 38). The median day of the cycle for iAFC determination was day 4 (range 2 – 5). The median day of the cycle for sAFC determination was day 9 (8 – 11). The median number of cycles between iAFC and sAFC was 1 (IQR 0 – 3).

There was a significant difference between median iAFC and median sAFC (16 AFC [IQR 9 – 24] vs. 13 AFC [IQR 7 – 21], p=0.001). There was moderate agreement between iAFC and sAFC for categorization into predicted low response, normal response and risk of high response (kappa = 0.525). 54/79 (68.4%, 95% CI 58.2 – 78.7) patients were allocated into the same categories by iAFC and sAFC (Table 1).

The iAFC and sAFC showed moderate CCC (0.678, 95% CI 0.543 – 0.779) with 32.2% variation in AFC within the same subject between different cycle phases. A Bland Altman plot showed a mean difference between measurements of 2 (95% CI 0.63 – 4.0) with LoA between -13 and 17. There was significant inequality of variance between women with <13 follicles and the rest (p=0.002) (Figure 3). The LoA in women with fewer follicles was -6 to 6 compared to -14 to 21 in those with more follicles.

In the blinded protocol design exercise there was good agreement between COS protocol types selected based on iAFC and those based on sAFC (kappa = 0.750). There was good agreement for starting gonadotrophin dose between COS
protocols designed based on iAFC and sAFC with a CCC of 0.970 (95% CI 0.951 – 0.982).

41/80 women underwent a stimulated IVF cycle. The median total hMG was 2700 iu/L (IQR 1462.5 – 4275) and the median number of oocytes retrieved was 12 (IQR 6 – 15). There was a significant positive correlation between iAFC, sAFC, AMH and the number of oocytes retrieved. 4/41 women (9.8%, 95% CI 0.7 – 18.9) had ≤4 oocytes retrieved and 7/41 (17.1%, 95% CI 5.6 – 28.6) had ≥16 oocytes retrieved. The AUC for prediction of poor response and risk of OHSS for age, FSH, AMH, iAFC and sAFC is shown in Table 2.

Discussion
Our study aimed to examine whether performing AFC determination in the late follicular phase would have an impact on the clinical management of infertile women. Our results showed that there is a statistically significant difference between AFCs determined in the early follicular and those determined in the late follicular phase which leads to poor agreement between them. However this poor agreement does not appear to have a significant impact on COS protocol design and the predictive value of late follicular phase AFC for extremes of ovarian response.

Our study is the first to prospectively examine AFC variation in a population of infertile women which lends validity to our findings. Previous studies have examined the intra cycle variation of AFC in healthy volunteers. Deb et al. performed a study on 38 women scanned at 4 different instances in the same cycle. They employed sonoAVC in a relatively young population (mean age 28 years) and found minimal AFC intra cycle variation, in the order of 6% (Deb et al. 2013). In contrast, van Disseldorp et al. in an older population of patients (median age 33 years) report an approximately 30% intra cycle AFC variation (van Disseldorp et al. 2010). This level of intra cycle variation led van Disseldorp et al. to conclude that restricting AFC to the early follicular phase would improve reliability of the test. Our results are similar to those by van Disseldrop et al. and
we confirmed that AFCs show significant variation reaching 30% between early and late follicular phases of the cycle.

In contrast to van Disseldopr et al. in our study the majority of paired AFC measurements were not performed in the same cycle but consecutive cycles but never more than 6 cycles apart. While this interval restriction ensures that the observed effect reflects inter and intra cycle variation rather than underlying age related decline, it is not possible to discern directly from our results whether the variation we observed is due to intra or inter cycle change. This would require examinations in the same cycle followed by examinations in a subsequent cycle which may be difficult to achieve. Others have examined the inter cycle variation of early follicular AFCs and both Elter et al. and Bancsi et al. showed that AFC exhibits significant inter cycle variability with the coverage interval of the difference ranging to a minimum of +/- 6 follicles (Bancsi et al. 2004; Elter et al. 2005). Compared to these authors we observed greater limits of agreement between paired AFCs which may be due to our population’s considerably higher median AFC. When we restricted the population to women with less than total 13 follicles the limits of agreement in our study were identical to those previous reports (+/- 6). This is despite the fact the determinations happened in different cycle times. From this it would appear that two AFCs determined in different cycles and different times in the cycle are not more variable than AFCs determined in different cycles at the same cycle time.

One of the purposes of biomarkers of ovarian response is to alert clinicians to the risk of poor ovarian response to stimulation and, at the other end of the spectrum, the risk of high response leading to OHSS. In this study we wanted to examine whether the degree of AFC variation we and others have observed would impact this categorisation. Indeed, we found only moderate agreement between iAFC and sAFC for classification into ovarian reserve categories. The majority of discrepancy appears in the classification into “at risk of high response” category where 11 patients considered “at risk of high response” with early follicular AFC would have been classified as “normal” by sAFC. It is questionable however whether restriction to the early follicular phase would
have addressed this issue as it may represent intercycle variability rather than intracycle variability and thus an inherent limitation of AFC. ORT is also used to design treatment protocols for ovarian stimulation during an IVF cycle. (Broer et al. 2013; Jayaprakasan et al. 2012). We showed here that using AFC determined in the late follicular phase for the purpose of designing a COS protocol would not have lead to a change in either type of protocol selected or starting gonadotrophin dose. Finally it appears that AFC timing does not impact the predictive ability of the test for extremes of response at the time of COS with both iAFC and sAFC showing a similar area under the curve for the prediction of poor response and risk of OHSS. This is consistent with the retrospective study by Rombauts et al. showing that the AFC retains its predictive ability irrespective of the time in the cycle it was determined (Rombauts et al. 2011).

AFC as a test of ovarian reserve has been criticized for substantial operator dependency (Iliodromiti et al. 2014). Nevertheless authors have shown that there is adequate interobserver agreement for AFC determination using both 3D and 2D transvaginal ultrasound (Scheffer et al. 1999). We aimed to produce a pragmatic study and so did not restrict the operators included. In order to assess the robustness of our results, we examined the reliability of AFC determination in our unit. In agreement with Sheffer et al. we demonstrated excellent ICC between operators included in this study suggesting that the AFC variation we observed represents a real difference rather than inter observer variability.

One of the reasons the current consensus restricts AFC timing to days 2 - 5 of the cycle is an attempt to standardize AFC measurements (Broekmans et al. 2010). However women dislike being scanned while menstruating and achieving this timing creates unnecessary administrative burden for the clinic. Transvaginal scanning provides a wealth of information beyond AFC and will remain an integral part of infertility assessment. However the greater cycle stability and operator independence of AMH suggest that AFC will become a complementary test rather than a main ORT determinant. We show here that AFCs determined in different cycle phases have significant statistical difference but similar clinical
validity. Based on our results it may be possible to combine AFC determination in the late follicular phase with other ultrasound based tests such as 3D SIS and HyCoSy without significant loss of clinically useful information.
References


van Disseldorp J, Lambalk CB, Kwee J, Looman CW, Eijkemans MJ, Fauser BC, Broekmans FJ. Comparison of inter- and intra-cycle variability of anti-
Table 1: Agreement between iAFC and sAFC for categorization into at risk of low response (total AFC ≤ 4), normal and at risk of high response (total AFC ≥ 16) (k = 0.525)

<table>
<thead>
<tr>
<th></th>
<th>Low response</th>
<th>Normal response</th>
<th>High response</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>iAFC</strong> Low response n (%)</td>
<td>2 (66.6)</td>
<td>1 (33.3)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td><strong>Normal response n (%)</strong></td>
<td>5 (13.9)</td>
<td>27 (75.0)</td>
<td>4 (11.1)</td>
<td>36</td>
</tr>
<tr>
<td><strong>High response n (%)</strong></td>
<td>0</td>
<td>11 (27.5)</td>
<td>29 (72.5)</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Area under the curve (95% CI)</td>
<td>Risk of OHSS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poor response</td>
<td>Risk of OHSS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>iAFC</td>
<td>0.932 (0.857 – 1.0)</td>
<td>0.701 (0.533 – 0.869)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sAFC</td>
<td>0.927 (0.850 – 1.0)</td>
<td>0.737 (0.504 – 0.971)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMH</td>
<td>0.880 (0.762 – 0.998)</td>
<td>0.691 (0.485 – 0.898)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSH</td>
<td>0.625 (0.334 – 0.917)</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.720 (0.509 – 0.931)</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Area under the curve for the prediction of poor response (oocytes ≤4) and risk of OHSS (oocytes ≥16) after COS using a variety of ovarian response biomarkers (n=41)
Bland Altman chart for interobserver agreement for determination of total antral follicle count (DM - AA)
Bland Altman chart for interobserver agreement for determination of total antral follicle count (DM - VT)
Bland Altman chart for agreement of total antral follicle count determined in early and late follicular phase of the cycle (iAFC - sAFC)
Dimitrios Mavrellos has been NIHR academic clinical lecturer in Reproductive Medicine at University College London since 2012. He studied medicine at the University of Oxford and Guy's King’s St Thomas’ School of Medicine. He completed his core Obstetrics and Gynecology training in London and undertook his MD in gynecological ultrasound. He is currently a subspecialty trainee in Reproductive Medicine at University College London Hospital. His research interests include the diagnosis and management of endometriosis, the diagnosis and management of early pregnancy complications, and the impact of gynecological disorders on assisted conception techniques.
We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Dr Dimitrios Mavrellos
Dear Mr Mavrelos,

Thank you for your email. I can confirm that from the form, your project is not research and therefore does not require ethical review by a NHS or Social Care Research Ethics Committee or management permission through the NHS R&D office. Please approach your clinical audit lead (attached find the list) who will be able to advise on the next steps.

Kind regards,

Tabitha Kavoi
Senior Portfolio Coordinator
Joint Research Office
(part of the Research Support Centre)
1st Floor Maple House (Suite B)
149 Tottenham Court Road
London W1T 7DN

Telephone: 020 3447 5696 Internal: 75696
Fax: 020 3108 2312
Postal Address: Joint Research Office, UCL, Gower Street, London WC1E 6BT
Web: www.ucl.ac.uk/jro

UCLH Research Patient Flag, live on e-CareLogic (CDR) from the 21st September 2015

-----Original Message-----
From: Mavrelos, Dimitrios [mailto:d.mavrelos@ucl.ac.uk]
Sent: 29 September 2015 15:48
To: RandD
Subject: Audit assessment form

Dear R&D,

Please find attached a form for a project we are hoping to undertake in the RMU. I will be grateful for your assessment.

Thank you

Dimitri

Mr Dimitrios Mavrelos
Clinical Lecturer in Reproductive Medicine University College Hospital d.mavrelos@nhs.uk
0044 7958384268

This email is confidential and is intended solely for the person or entity to whom it is addressed. If this is not you, please forward the message to mail.administrator@uch.nhs.uk. We have scanned this email before sending it, but cannot guarantee that malicious software is absent and we shall carry no liability in this regard.

We advise that information intended to be kept confidential should not be sent by email. We also advise that health concerns should be discussed with a medical professional in person or by telephone. NHS 111 can also provide advice. We shall not be liable for any failure to follow this advice. University College London Hospitals NHS Foundation Trust (UCLH).
We sought the opinion of the Joint Research Office of the hospital and were advised that formal ethics approval was not required as the project involved no change in routine clinical practice. Please find attached the relevant correspondence.

Dr DImitrios Mavrelos