Comparison of antimüllerian hormone levels and antral follicle count as predictor of ovarian response to controlled ovarian stimulation in good-prognosis patients at individual fertility clinics in two multicenter trials

Scott M. Nelson, M.R.C.O.G., Ph.D.,[a] Bjarke M. Klein, Ph.D.,[b] and Joan-Carles Arce, M.D., Ph.D.[c]

[a] School of Medicine, University of Glasgow, Glasgow, United Kingdom; and [b] Global Biometrics and [c] Reproductive Health, Ferring Pharmaceuticals, Copenhagen, Denmark

Objective: To compare antimüllerian hormone (AMH) and antral follicle count (AFC) as predictors of ovarian response to controlled ovarian stimulation at individual fertility clinics.
Design: Retrospective analysis of individual study center data in two multicenter trials. Centers that provided >10 patients were included in the analysis.
Setting: A total of 19 (n = 519 patients) and 18 study centers (n = 686 patients) participating in a long GnRH agonist trial (MERIT) and a GnRH antagonist trial (MEGASET), respectively.
Patient(s): Infertile women of good prognosis.
Intervention(s): Long GnRH agonist or GnRH antagonist cycles.
Main Outcome Measure(s): Correlation between AMH and AFC, and oocyte yield by each study center for each trial.
Results(s): Antimüllerian hormone was more strongly correlated with oocyte yield than AFC: $r = 0.56$ vs. $r = 0.28$ in the GnRH agonist cohort, and $r = 0.55$ vs. $r = 0.33$ in the GnRH antagonist cohort. The correlation was numerically higher for AMH than for AFC at a significantly higher proportion of study centers: 17 (89%) and 15 (83%) centers in the long GnRH agonist and GnRH antagonist trial, respectively. Assessment of the relative capacity of AMH and AFC for predicting oocyte yield demonstrated that AMH dominated the model: AMH, $R^2 = 0.29$ and 0.23; AFC: $R^2 = 0.07$ and 0.07; AMH + AFC: $R^2 = 0.30$ and 0.23 for long GnRH agonist and GnRH antagonist trials, respectively.
Conclusions(s): Antimüllerian hormone was a stronger predictor of ovarian response to gonadotropin therapy than AFC at the study center level in both randomized trials utilizing GnRH agonist and GnRH antagonist protocols. Antral follicle count provided no added predictive value beyond AMH. (Fertil Steril® 2015;103: 923–30. ©2015 by American Society for Reproductive Medicine.)
Key Words: Antimüllerian hormone, antral follicle count, IVF clinic, multicenter randomized controlled trial, ovarian response

Discuss: You can discuss this article with its authors and with other ASRM members at http://fertstertforum.com/nelsons-amh-afc-controlled-ovarian-stimulation/

Use your smartphone to scan this QR code and connect to the discussion forum for this article now.*

* Download a free QR code scanner by searching for "QR scanner" in your smartphone's app store or app marketplace.

Received October 10, 2014; revised December 10, 2014; accepted December 15, 2014; published online January 24, 2015.
S.M.N. is a consultant for Beckman Coulter, Roche, Ferring, Merck Serono, and MSD; received speaker fees from Merck Serono, Ferring and Beckman Coulter; and has grants from the Wellcome Trust, Medical Research Council (UK), UK Clinical Research Collaboration, National Institutes of Health, and Chief Scientist Office (Scotland). B.M.K. has nothing to disclose. J.-C.A. has nothing to disclose.
Reprint requests: Joan-Carles Arce, M.D., Ph.D., Ferring Pharmaceuticals A/S, Reproductive Health, Global Clinical & Non-Clinical R&D, Kay Fiskers Plads 11, DK-2300 Copenhagen S, Denmark (E-mail: jca@ferring.com).

Fertility and Sterility® Vol. 103, No. 4, April 2015 0015-0282

http://dx.doi.org/10.1016/j.fertnstert.2014.12.114

The ovarian response resulting from controlled ovarian stimulation (COS) in IVF with standard doses of gonadotropins is associated with a large interindividual variability. Individualization of the starting dose of gonadotropin according to ovarian reserve parameters has been proposed as a means of improving safety and efficacy of COS (1–3). To date, a number

of markers of ovarian response have been used and evaluated (4), such as age, FSH, and inhibin B; however, antral follicle count (AFC) and antimüllerian hormone (AMH) are the two biomarkers that have consistently provided the best performance in terms of predicting ovarian response to gonadotropins (5, 6).

Antral follicle count has been shown to possess similar performance as AMH in predicting the number of oocytes retrieved in the majority of single-center observational cohort studies in IVF/intracytoplasmic sperm injection (ICSI) patients treated with GnRH agonist protocols (7–11), whereas a few studies have suggested either AFC (12, 13) or AMH (14, 15) as being a better predictor. Two meta-analyses of a number of these relatively small, single-center studies indicated that AMH and AFC have similar levels of accuracy and clinical value for the prediction of poor (16) as well as excessive response (17). In marked contrast to these reports, three recent large, prospective, multicenter trials in IVF/ICSI patients of good prognosis consistently concluded that AMH was a better predictor of ovarian response than AFC in GnRH agonist (18) and antagonist (19, 20) cycles, regarding the number of oocytes retrieved as well as categorization of low and high responders. Furthermore, in models of ovarian response AFC did not provide any additional predictive value beyond that provided by AMH (18–20).

The overall superior performance of AMH over AFC in these multicenter trials may have been attributed to considerable sonographer-dependent variability across centers. Furthermore, such interoperator variability between different IVF clinics may also explain the different performance of AFC in single-center and multicenter studies. Therefore, it would be important to explore whether the findings in multicenter trials are determined by the integrated data evaluation rather than by the actual performance at each study center, because AMH was analyzed centrally and AFC locally. This is essential to clarify, given the perception that each fertility clinic believes that their ultrasound evaluation of AFC is robust, and because AFC has been considered the gold standard biomarker for the prediction of ovarian response.

The aim of the present study was to compare the values of AMH and AFC for prediction of oocyte yield at a study center level for fertility clinics participating in two large, multicenter trials: one conducted with the long GnRH agonist protocol (21) and the other with a GnRH antagonist protocol (22). Consistent with previous systematic reviews and individual patient data meta-analyses (4, 23, 24), other predictors of ovarian response to gonadotropin stimulation, such as FSH, were shown to be less predictive than AMH in both trials (18, 19) and therefore not considered for this evaluation.

MATERIALS AND METHODS

Study Population and Study Centers

This study is a retrospective analysis of data prospectively collected in two randomized, controlled, multicenter trials in IVF/ICSI patients of good prognosis undergoing COS with highly purified hMG (Menopur; Ferring Pharmaceuticals) or recombinant FSH (follitropin alfa [Gonal-F, Merck Serono] or follitropin beta [Puregon, MSD]) after a long GnRH agonist protocol (MERIT trial) (21) or a GnRH antagonist protocol (MEGASET trial) (22). The women included in each trial had been infertile for at least 1 year and had a regular menstrual cycle, a transvaginal ultrasound documenting presence and adequate visualization of both ovaries without evidence of abnormality, and an early follicular-phase serum level of FSH within normal limits (1–12 IU/L). Women with polycystic ovary syndrome and/or a poor response in a previous COS cycle were excluded in both trials. In the GnRH antagonist trial, women with an AFC <10 (diameter 2–10 mm) at screening were excluded. At each study center, the patients were randomized in a 1:1 ratio to treatment with either highly purified hMG or rFSH.

The trials were conducted in accordance with the Declaration of Helsinki, the International Conference on Harmonization Guidelines for Good Clinical Practice, and local regulatory requirements, and were approved by the local regulatory authorities and the independent ethics committees covering all participating study centers. Written informed consent was provided by each of the subjects.

Stimulation Regimens

In the long GnRH agonist trial, pituitary suppression was initiated with 0.1 mg/d of triptorelin (Decapeptyl, Ferring Pharmaceuticals) 5–7 days before the estimated start of next menses and continued until the end of gonadotropin administration. Gonadotropin treatment started when down-regulation was confirmed, and the dose was fixed at 225 IU/d for the first 5 days of COS, followed by individual dose adjustments according to the patient's follicular response. In the GnRH antagonist trial, treatment with a daily dose of 150 IU of gonadotropin started on day 2–3 of the menstrual cycle and was fixed for the first 5 days of COS, followed by individual dose adjustments according to the patient's follicular response. Treatment with 0.25 mg/d of ganirelix (Orgalutran, MSD) was initiated on stimulation day 6 and continued throughout the gonadotropin treatment period. In both trials, the criteria for giving hCG (Ovitrelle, Merck Serono) was development of at least three follicles with a diameter ≥ 17 mm. Oocyte retrieval took place 36 ± 2 hours later, followed by IVF or ICSI and embryo/blastocyst transfer. Detailed descriptions of ovarian stimulation regimens, cohort assessments, and procedures in the long GnRH agonist and antagonist trials are provided in Nyboe Andersen et al. (2006) (21) and Devroey et al. (2012) (22), respectively.

Endocrine Assays and Antral Follicle Count

In both trials, circulating concentrations of AMH were analyzed in serum samples collected on stimulation day 1 before the start of stimulation by a central laboratory (Hormone Laboratory, Universitair Ziekenhuis, Brussels, Germany for the agonist trial and Laboratory for Clinical Research, Kiel, Germany for the antagonist trial). Serum samples were immediately frozen to $-18^{\circ}C$ for the first 2 weeks until transport to a central facility, followed by storage at $-70^{\circ}C$. Antimüllerian hormone was analyzed by

an enzyme-linked immunosorbent assay (long GnRH agonist trial: Immunotech Beckman Coulter AMH ELISA [A11893]; GnRH antagonist trial: Beckman Coulter Gen 2 AMH ELISA [A79765]; 1 ng/mL = 7.143 pmol/L) without predilution. The AMH assays had a sensitivity of 0.35 and 0.08 ng/mL and total imprecision (percent coefficient of variation) of <9.5 and <7.7 in Immunotech Beckman Coulter and Beckman Coulter Gen 2, respectively. Antral follicle count was the total number of follicles with a diameter between 2 and 10 mm in both ovaries on stimulation day 1 before the start of stimulation, as measured by transvaginal ultrasound by local sonographers at each study center.

Statistical Analysis

Selected baseline characteristics and outcome variables are presented using mean and SD, or as frequencies and percentages. The populations were compared using *t* tests and χ^2 tests. The Spearman rank correlation coefficients between oocytes retrieved and the potential predictor variables "AMH" and "AFC" were calculated within each individual study center, and within each study cohort. The capability of AMH and AFC to predict the number of oocytes retrieved was evaluated using a stepwise forward selection procedure within an analysis of covariance model framework. The procedure sequentially selected the predictor variables according to the increase in the coefficient of determination (R^2). The relative importance of AMH and AFC with respect to predicting the number of oocytes retrieved was further illustrated using partial residual plots. Besides the predictor variables AMH and AFC, the variable "Study Center" was also included as a factor accounting for differences in oocyte pickup procedures/effectiveness between the clinics. Study centers included in the primary analysis were those providing more than 10 patients who underwent oocyte retrieval in either trial and had assessments of both AMH and AFC before start of stimulation. This cut-off was a compromise between not including centers with too few patients, which could potentially give misleading results in the evaluation of individual study centers, and not excluding too many study centers/patients. Two additional sensitivity analyses were performed for the forward selection procedure: inclusion of all contributing centers, and restriction to those centers contributing ≥20 patients. All statistical analyses were performed in SAS version 9.3 (SAS Institute).

RESULTS

In total, the long GnRH agonist trial and the GnRH antagonist trial comprised 37 and 25 study centers, respectively. Of these, 19 study centers in the long GnRH agonist trial and 18 centers in the GnRH antagonist trial contributed more than 10 patients who underwent oocyte retrieval and had baseline measurements of AMH and AFC (Supplemental Fig. 1, available online). Demographics, baseline characteristics, and main outcome parameters of the two study cohorts (long GnRH agonist trial: n = 519 patients; GnRH antagonist trial: 686 patients) are presented in Table 1. There were no clinically relevant differences between the two cohorts regarding the demographics and serum levels of AMH on stimulation day 1. At start of COS, the down-regulated patients in the long GnRH agonist cohort had lower AFC, FSH, LH, E_2, and P than the patients in the GnRH antagonist cohort. Regarding stimulation and treatment outcome, the mean number of oocytes retrieved was higher in patients treated with the long GnRH agonist protocol compared with that in the GnRH antagonist protocol, but the observed live birth rates of the fresh stimulation cycle were similar (i.e., 27% and 28% in the long GnRH agonist and GnRH antagonist cohorts, respectively).

TABLE 1

Demographics, baseline characteristics and outcome.

Variable	Long GnRH agonist trial (19 study centers, 519 patients)	GnRH antagonist trial (18 study centers, 686 patients)	*P* value
Demographics			
Age (y)	30.9 ± 3.2	30.5 ± 2.7	.059
Body weight (kg)	61.9 ± 8.3	60.2 ± 6.9	<.001
BMI (kg/m^2)	22.2 ± 2.6	22.0 ± 2.0	.103
Cycle length (d)	28.4 ± 1.7	28.6 ± 1.9	.022
Duration of infertility (y)	4.0 ± 2.3	3.2 ± 1.8	<.001
First treatment cycle	365 (70)	513 (75)	.085
Stimulation day 1			
AMH (pmol/L)	29.8 ± 16.5	27.3 ± 19.1	.013
AFC (n)	11.3 ± 6.4	15.7 ± 5.6	<.001
FSH (IU/L)	4.0 ± 2.1	7.4 ± 2.1	<.001
LH (IU/L)	2.3 ± 1.4	6.2 ± 2.2	<.001
E_2 (pmol/L)	42.3 ± 23.5	178 ± 94	<.001
P (nmol/L)	1.2 ± 0.6	2.3 ± 2.1	<.001
Outcome			
Oocytes retrieved (n)	11.0 ± 5.7	9.9 ± 5.6	<.001
Ongoing pregnancy rate	141 (27)	199 (29)	.482
Live birth rate	140 (27)	193 (28)	.656

Note: Values are mean ± SD or n (%).

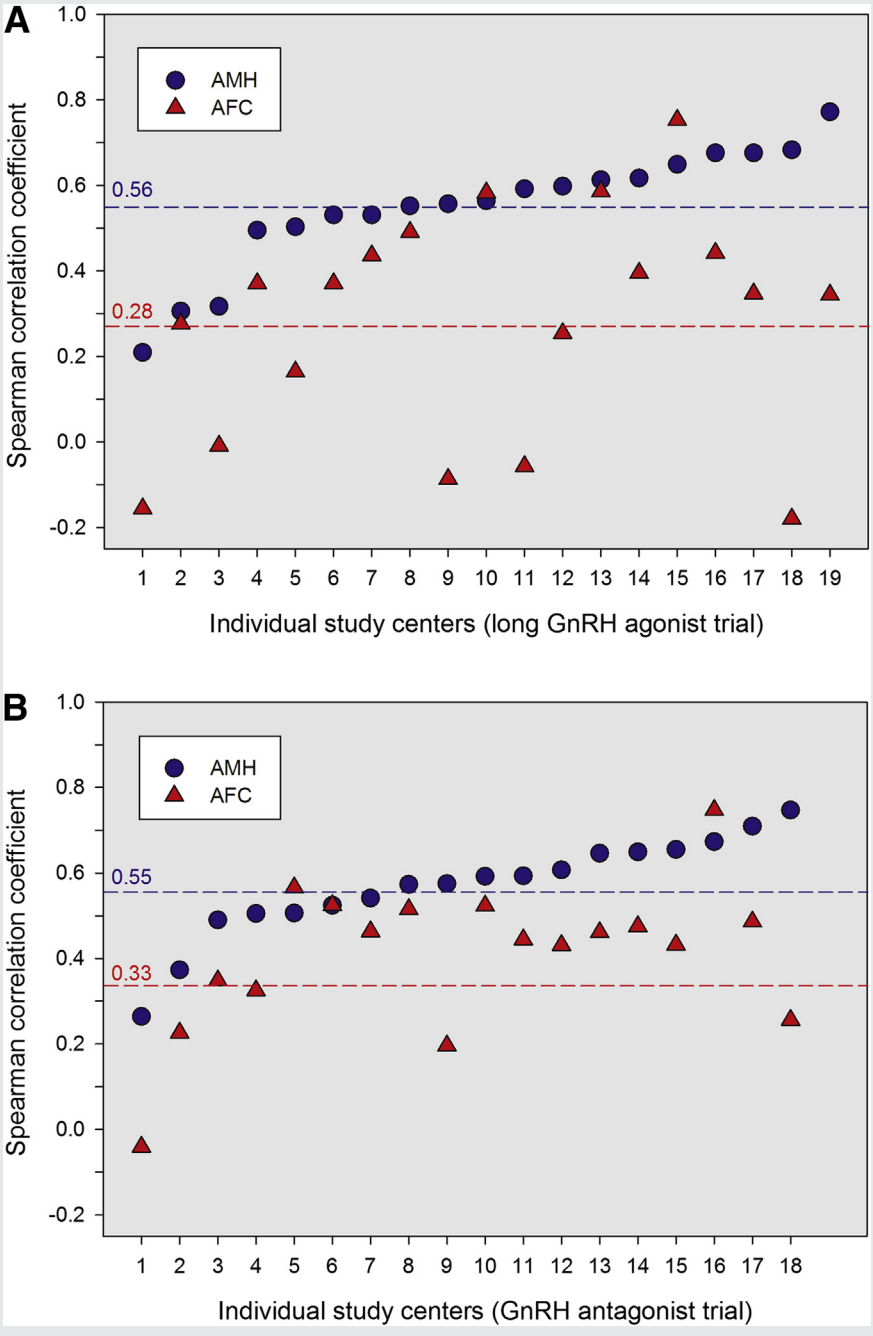
Nelson. AMH vs. AFC as predictor of oocyte yield. Fertil Steril 2015.

AMH and AFC as Biomarkers of Ovarian Response at Individual IVF Clinics

Overall, AMH was a markedly better predictor of the number of oocytes retrieved than AFC in both study cohorts. The Spearman correlation coefficient was 0.56 (95% confidence interval [CI] 0.50–0.62) for AMH in the long GnRH agonist trial and 0.55 (95% CI 0.50–0.60) in the GnRH antagonist trial, as compared with 0.28 (95% CI 0.20–0.36) and 0.33 (95% CI 0.26–0.39), respectively, for AFC (Fig. 1). Furthermore, in patients participating in the GnRH antagonist trial where ICSI was used as a method of insemination, the correlation between AMH and number of metaphase II oocytes was higher than that between AFC and metaphase II oocytes (0.45 [95% CI 0.39–0.51] vs. 0.30 [95% CI 0.23–0.37]). In addition, sensitivity analyses showed that AMH had higher correlation with oocyte yield than AFC, irrespective of whether women achieved a pregnancy (data not shown).

At individual IVF clinics, the correlation coefficient with oocyte yield for AMH was numerically higher than that for AFC in 17 (89%) of the 19 study centers in the long GnRH agonist trial and in 15 (83%) of the 18 study centers in the GnRH antagonist trial (Fig. 1). In both trials, the observed

FIGURE 1



Correlations between AMH and AFC, respectively, and number of oocytes retrieved in patients participating in a long GnRH agonist trial (**A**) and a GnRH antagonist trial (**B**) at individual IVF clinics. The dotted lines show the overall mean correlation coefficients of AMH and AFC for the study cohorts. Antimüllerian hormone was a stronger predictor of oocyte yield (i.e., a difference in correlation coefficient of >0.10) in study center nos. 1, 3, 4, 5, 6, 9, 11, 12, 14, 16, 17, 18, and 19 in the long GnRH agonist trial, and in study center nos. 1, 2, 3, 4, 9, 11, 12, 13, 14, 15, 17, and 18 in the GnRH antagonist trial. Only study center no. 15 in the long GnRH agonist trial exhibited a stronger association with AFC.

Nelson. AMH vs. AFC as predictor of oocyte yield. Fertil Steril 2015.

outcome (i.e. that AMH was better than AFC at most centers) was significantly different compared with the outcome expected if AMH and AFC were equally good predictors (binomial test: $P<.001$ and $P=.007$ in the long GnRH agonist and GnRH antagonist trials, respectively). When considering a difference in correlation coefficient of ≤ 0.10 (the smallest width of the 95% confidence interval in the analysis) between AMH and AFC as an indicator of similar performance, AMH was a stronger predictor of oocyte yield in 13 study centers (68%) in the long GnRH agonist trial and 12 study centers (67%) in the GnRH antagonist trial. Only one center (5%) exhibited stronger association with AFC in the long GnRH agonist trial.

Stepwise Regression Model

The stepwise forward procedure for prediction of number of oocytes retrieved provided similar findings on the contribution of the selected predictors for the two cohorts (Table 2). Antimüllerian hormone was identified as the variable with the highest coefficient of determination: $R^2 = 0.29$ and 0.23 for the long GnRH agonist trial and the GnRH antagonist trial, respectively. The R^2 for AFC was only 0.07 in both trials, and inclusion of AFC in the models had no significant improvement on the prediction of oocyte yield. In the long GnRH agonist trial, the R^2 increased slightly, from 0.29 (AMH) to 0.30 (AMH + AFC), whereas there was no change in the GnRH antagonist trial (i.e., $R^2 = 0.23$ both for AMH and AMH + AFC). Similar to the results obtained in the two study cohorts, sensitivity analyses showed comparable outcomes when restricting the analysis to centers with ≥ 20 patients (long GnRH agonist trial [n = 11 centers]: $R^2 = 0.28$ and 0.07 for AMH and AFC, respectively; GnRH antagonist trial [n = 11 centers]: $R^2 = 0.23$ and 0.06 for AMH and AFC, respectively), as well as when including all study centers with AMH and AFC data (long GnRH agonist trial: $R^2 = 0.28$ and 0.07 for AMH and AFC, respectively; GnRH antagonist trial: $R^2 = 0.24$ and 0.08 for AMH and AFC, respectively).

The relative importance of AMH and AFC with respect to predicting number of oocytes retrieved is further illustrated in Figure 2. The plots of partial residuals obtained after adjusting the number of oocytes retrieved for AFC and Study Center against AMH indicated a strong linear relationship (Fig. 2A and B). Reversing the roles of AMH and AFC indicated that only a small improvement is gained when adding AFC if AMH is already included in the model (Fig. 2C and D).

In a hypothetical prediction model that would consider study site contribution for explaining oocytes retrieved, inclusion of the variable Study Center would lead to the second largest improvement of the coefficient of determination besides AMH (i.e., from 0.29 to 0.40 and from 0.23 to 0.33 for the long GnRH agonist trial and the GnRH antagonist trial, respectively) (Table 2). Also in this case, inclusion of AFC did not improve the prediction of oocyte yield, either in the long GnRH agonist or the GnRH antagonist cohort.

DISCUSSION

At present, several single-center studies have suggested that AMH and AFC have the same level of accuracy and clinical value for ovarian response prediction in women undergoing IVF treatment (5, 16, 17). The assumption that these two biomarkers of ovarian reserve possess similar performance has, however, been questioned by three recent large, multicenter trials (18–20). In each of these trials, analysis of the integrated data obtained from the participating study centers showed that AFC by itself was a poorer predictor of the ovarian response to COS than AMH, and furthermore, that AFC provided no added predictive value beyond AMH.

The finding that AMH was a more robust biomarker of the ovarian response to gonadotropins across multicenter trials than AFC was also confirmed in the present study, which evaluated individual study center data from two of these trials, one using a long GnRH agonist protocol (21) and the other a GnRH antagonist protocol (22). The correlation coefficient for AMH and number of oocytes retrieved was numerically higher in most centers (approximately 85%), regardless of the protocol and the different gonadotropin doses used, because stimulation was started with a dose of 225 IU in the long GnRH agonist protocol and 150 IU in the GnRH antagonist protocol. The type of protocol or the different starting doses between trials may have influenced the oocyte yield, because the mean number of oocytes retrieved was slightly lower in the GnRH antagonist trial than in the long GnRH agonist trial. Nevertheless, because the treatment effect was constant for both AMH and AFC, it would not be expected to alter the strength of association for the two biomarkers. It should be noted that AFC was measured on stimulation day 1 in both trials. However, the timing of ultrasound assessment in relation to the pituitary desensitization in the long agonist protocol is not considered to have significantly changed the AFC, because several previous studies have demonstrated that the AFC is unaffected by down-regulation (25–28). Furthermore, AFC has been shown to possess similar value for prediction of ovarian response irrespective of whether the assessment is performed before or after pituitary down-regulation (26).

In comparison to AFC, the correlation coefficient for AMH and number of oocytes retrieved was remarkably constant across centers. The observed variability of the correlation coefficients for AFC suggests not only a marked difference in the performance of the ultrasound measurements by the operators at the different clinics, it also indicates a potential variability in AFC within clinics, because two-thirds of the study centers in both trials had considerably lower correlation between their AFC measurements and the ovarian response compared with the serum assay of AMH. Furthermore, inclusion of AFC in stepwise logistic regression models did not improve the prediction of oocyte yield, consistent with previous analyses using integrated data from all participating centers (18–20). The somewhat lower apparent variability of AFC values in the GnRH antagonist trial compared with the long GnRH agonist trial may be explained by the time interval between the two trials (2009/2010 and 2004, respectively), because ultrasound technology has improved during these years (29). However, it should be noted that the value of AFC as a predictor of oocyte yield was not improved in the GnRH antagonist trial.

Despite the timing of AFC being standardized for both multicenter trials, biological and technical explanations

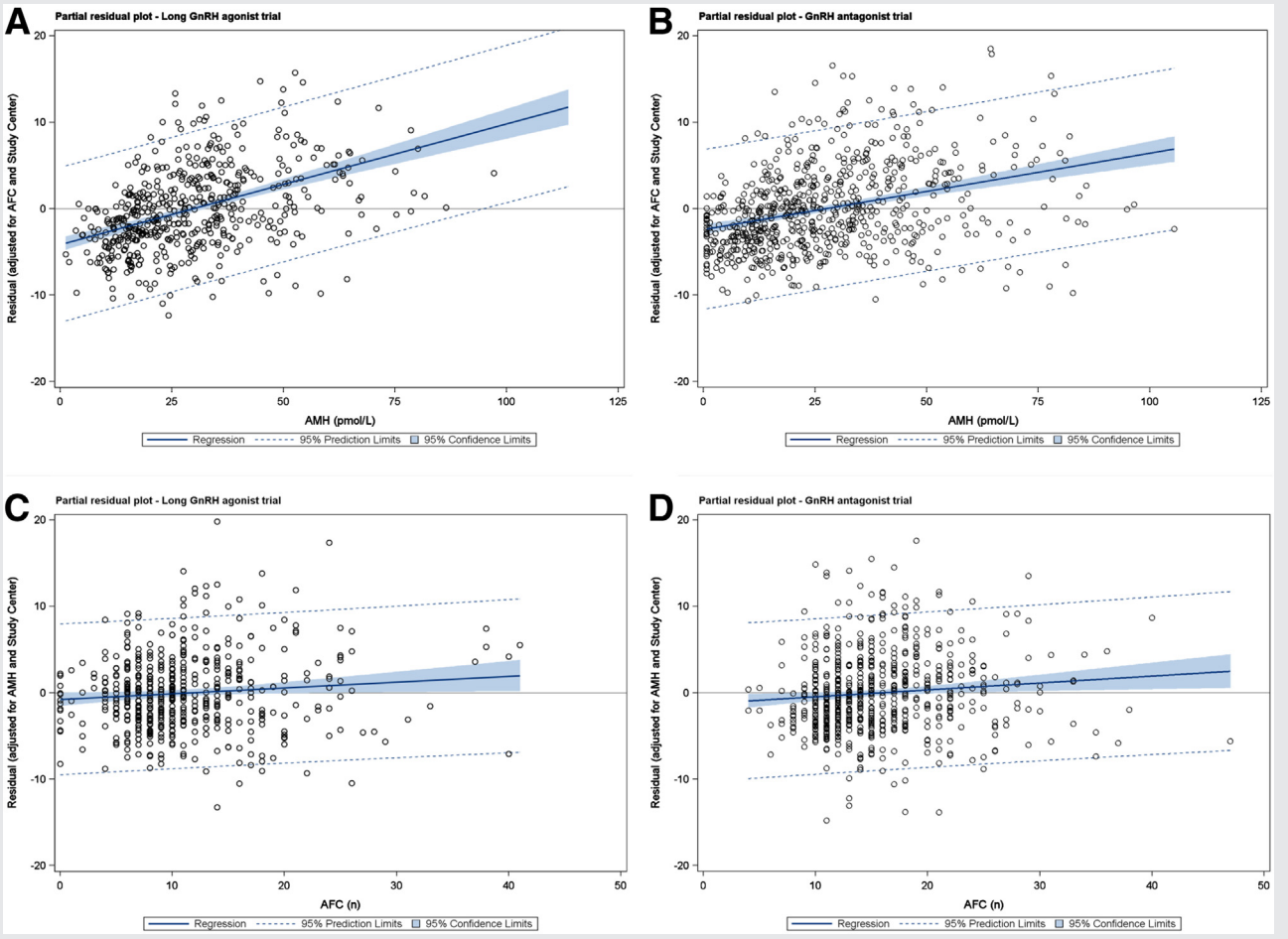
TABLE 2

Stepwise analysis of covariance models.

Model	R^2, long GnRH agonist trial	R^2, GnRH antagonist trial
AMH	0.29	0.23
AFC	0.07	0.07
Study Center	0.11	0.07
AMH + AFC	0.30	0.23
AMH + Study Center	0.40	0.33
AMH + Study Center + AFC	0.41	0.34

Nelson. AMH vs. AFC as predictor of oocyte yield. Fertil Steril 2015.

FIGURE 2



Panels illustrating the capability of AMH and AFC to predict the number of oocytes retrieved for the long GnRH agonist trial (*left*) and the GnRH antagonist trial (*right*). The upper panels (**A, B**) show the partial residuals from the model including study cohort, AFC, and Study Center plotted against AMH. The lower panels (**C, D**) show the partial residuals from the model including study cohort, AMH, and Study Center plotted against AFC.

Nelson. AMH vs. AFC as predictor of oocyte yield. Fertil Steril 2015.

may have contributed to its observed lower inter- and intra-clinic prediction performance (30). Ultrasound measurement of the functional ovarian reserve may overestimate the number of follicles that will be sensitive to gonadotropin stimulation because of possible inclusion of atretic follicles in the total count (31). In clinical practice, the measurement of AFC is known to show substantial intra- and interoperator variability regarding scanning techniques and methodology for counting and measuring follicles (25, 32), and larger clinics may work in teams with different sonographers. Furthermore, technical aspects of ultrasound equipment (e.g., resolution, depth) and different devices (two-dimensional vs. three-dimensional) may affect the performance of AFC measurements (33). Overall, the lack of reproducibility would emphasize the necessity for individual clinics to better standardize the assessment of AFC (34) but also to develop their own thresholds for prediction of ovarian response categories for AFC, rather than relying on existing literature.

Antimüllerian hormone was analyzed in a central laboratory, which may have reduced variation and improved overall performance, but this reflects current clinical practice for the majority of clinics. It should be noted that different AMH assays were used in the two multicenter trials (long GnRH agonist trial: Immunotech Beckman Coulter AMH ELISA; GnRH antagonist trial: Beckman Coulter Gen 2 AMH ELISA), and that both assays, like the AFC measurements, were associated with intra- and interindividual imprecision. Moreover, although the blood samples were managed and analyzed according to the manufacturers' instructions at the time of the trials, there was a potential risk of complement interference with the Beckman Coulter Gen 2 AMH assay (35). Nevertheless, despite these potential sources of errors regarding the assays, AMH performed markedly better than AFC in both trials. Antimüllerian hormone performance may also improve further with the introduction of assays in an automated platform with high reproducibility and not subject to complement interference (36).

The value of AMH and AFC for prediction of oocyte yield will also be affected by the oocyte retrieval procedure/operator at the IVF clinic, as indicated by the present prediction model. However, because AMH and AFC are highly linked to each other, it is anticipated that this center-specific effect, in other words, the ability of the center to accurately retrieve the oocytes available after stimulation, would affect equally the relationship between oocytes retrieved and AFC, as well as between oocytes retrieved and AMH levels. Nevertheless, the prediction model that accounted for differences in oocyte retrieval procedures between the study centers showed that AMH was the variable with the highest predictive value and that no further enhancement of the prediction of ovarian response could be achieved by the inclusion of AFC compared with AMH alone.

As mentioned above, the usefulness of AFC as a predictor of ovarian response has primarily been demonstrated in single-center, observational cohorts. Although the aim of an observational study is to examine the association of a predictor/exposure (AFC or AMH) with an outcome (oocyte yield), a major limitation is the potential for confounding (37, 38). Confounding may be reduced in observational studies by prevention in the design phase by restriction or matching; and adjustment in the statistical analysis by techniques such as stratification or multivariable analyses. These methods, however, require that the confounding variables are known and measured. Notably, few of the single-center studies that compared AMH and AFC as predictive biomarkers have undertaken this level of detailed analyses. It is therefore possible that in the observational cohort studies reported to date, confounding could have potentially strengthened or weakened the predictive effect of AFC and AMH against oocyte yield (39). Observational cohort studies may also be subject to selection bias, particularly because the value of the test may have influenced in advance the allocation of treatment and thus the outcome of interest (ovarian response). This was not the case in the present setting, in which all subjects were managed in a harmonized and similar manner independently of the variables of interest, including randomly assigned choice of gonadotropin preparation, similar schedule for monitoring of the follicular development, and a common triggering criterion for follicular maturation and same hCG dose. Although the generalizability of randomized controlled trials (RCTs) can be limited owing to the stricter inclusion criteria and rigid protocols to generate more homogenous study cohorts, RCTs are specifically designed to overcome the issues of differential confounding and selection bias between the treatment groups, making them strong candidates to examine the strength of association between exposures and outcomes of interest. Furthermore, in prediction modeling of ovarian response to COS with gonadotropins, the treatment effect may either be dependent on predictive factors or be constant. Although we are accustomed in reproductive medicine for the treatment effect to be modified by predictive factors (e.g., age), ideally there should be a constant treatment effect. This implies that we would require examination of how these predictors have performed in RCTs to truly assess their performance characteristics, where the treatment has been randomly allocated.

Although the present study has a number of strengths, including its size, the use of two prospective, multicenter RCTs utilizing two different stimulation strategies, inclusion of a large number of clinics with experience in clinical research, and robust statistical analyses, we do acknowledge several limitations. Only patients with an anticipated good prognosis to gonadotropin stimulation, based on serum FSH 1–12 IU/L (and AFC $\geq$10 in the GnRH antagonist trial), were included. Because expected poor responders were excluded, this selection may have attenuated the overall strength of the correlations given that only women within the normal range of AFC values were examined, but this limitation would also apply to AMH. The higher mean AFC value in the GnRH antagonist cohort compared with the long GnRH agonist cohort may be explained by more refined ultrasound equipment during the recent decade, and that AFC $\geq$10 was one of the inclusion criteria in this trial. Different machines were used for measurement of AFC at each of the centers, but this reflects current clinical practice, and center-specific differences were accounted for in the multivariate analysis, with no additional contribution from AFC observed. No data on individual sonographers were available, and it was therefore not possible to examine their contribution at a study center level. Finally, two sensitivity analyses, one after a further restriction to centers with $\geq$20 patients and the other after expanding the analysis cohorts to also include the centers contributing with $\leq$10 patients, resulted in an identical conclusion that AMH is a stronger predictor of oocyte yield than AFC. Although AMH was a poor predictor of embryo/blastocyst quality and ongoing pregnancy in the fresh cycle in both multicenter trials (18, 19), a positive association between AMH and cumulative live-birth rates in fresh and cryopreserved cycles has previously been reported in the GnRH antagonist trial. This is consistent with recent analyses of the strength of the association of AMH and live birth (40), and potentially reflects the availability of more oocytes and blastocysts for transfer in patients with higher AMH rather than a direct association between AMH and blastocyst quality (19).

In conclusion, when evaluating the data from two large, multicenter trials at the study center level, the analysis showed that AMH was a stronger predictor of ovarian response to gonadotropin therapy than AFC within the majority of the individual IVF clinics in both long GnRH agonist and GnRH antagonist protocols. Further, inclusion of AFC in the prediction models provided no added predictive value beyond AMH. These findings argue against the general assumption that the overall superior performance of AMH over AFC in multicenter trials is only attributed to marked sonographer-dependent variability across centers.

Acknowledgments: The authors thank Göran Pettersson, Ph.D., Reproductive Health, Ferring Pharmaceuticals, for assistance in writing the article.

REFERENCES

1. La Marca A, Sunkara SK. Individualization of controlled ovarian stimulation in IVF using ovarian reserve markers: from theory to practice. Hum Reprod Update 2014;20:124–40.

2. Nelson SM, Yates RW, Lyall H, Jamieson M, Traynor I, Gaudoin M, et al. Anti-müllerian hormone-based approach to controlled ovarian stimulation for assisted conception. Hum Reprod 2009;24:867–75.
3. Yates AP, Rustamov O, Roberts SA, Lim HY, Pemberton PW, Smith A, et al. Antimüllerian hormone-tailored stimulation protocols improve outcomes whilst reducing adverse effects and costs of IVF. Hum Reprod 2011;26: 2353–62.
4. 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.
5. La Marca A, Sighinolfi G, Radi D, Argento C, Baraldi E, Artenisio AC, et al. Antimüllerian hormone (AMH) as a predictive marker in assisted reproductive technology (ART). Hum Reprod Update 2010;16:113–30.
6. Nelson SM. Biomarkers of ovarian response: current and future applications. Fertil Steril 2013;99:963–9.
7. Elgindy EA, El-Haieg DO, El-Sebaey A. Antimüllerian hormone: correlation of early follicular, ovulatory and midluteal levels with ovarian response and cycle outcome in intracytoplasmic sperm injection patients. Fertil Steril 2008; 89:1670–6.
8. Jayaprakasan K, Campbell B, Hopkisson J, Johnson I, Raine-Fenning N. A prospective, comparative analysis of antimüllerian hormone, inhibin-B, and three-dimensional ultrasound determinants of ovarian reserve in the prediction of poor response to controlled ovarian stimulation. Fertil Steril 2010;93:855–64.
9. Lekamge DN, Barry M, Kolo M, Lane M, Gilchrist RB, Tremellen KP. Anti-müllerian hormone as a predictor of IVF outcome. Reprod Biomed Online 2007;14:602–10.
10. Muttukrishna S, McGarrigle H, Wakim R, Khadum I, Ranieri DM, Serhal P. Antral follicle count, antimüllerian hormone and inhibin B: predictors of ovarian response in assisted reproductive technology? BJOG 2005;112: 1384–90.
11. Van Rooij IA, Broekmans FJ, te Velde ER, Fauser BC, Bancsi LF, Jong FH, et al. Serum antimüllerian hormone levels: a novel measure of ovarian reserve. Hum Reprod 2002;17:3065–71.
12. Eldar-Geva T, Ben-Chetrit A, Spitz IM, Rabinowitz R, Markowitz E, Mimoni T, et al. Dynamic assays of inhibin B, antimüllerian hormone and estradiol following FSH stimulation and ovarian ultrasonography as predictors of IVF outcome. Hum Reprod 2005;20:3178–83.
13. Kwee J, Schats R, McDonnell J, Themmen A, de Jong F, Lambalk C. Evaluation of antimüllerian hormone as a test for the prediction of ovarian reserve. Fertil Steril 2008;90:737–43.
14. Ficicioglu C, Kutlu T, Baglam E, Bakacak Z. Early follicular antimüllerian hormone as an indicator of ovarian reserve. Fertil Steril 2006;85:592–6.
15. McIlveen M, Skull JD, Ledger WL. Evaluation of the utility of multiple endocrine and ultrasound measures of ovarian reserve in the prediction of cycle cancellation in a high-risk IVF population. Hum Reprod 2007;22: 778–85.
16. Broer SL, Mol BW, Hendriks D, Broekmans FJ. The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. Fertil Steril 2009;91:705–14.
17. Broer SL, Dólleman M, Opmeer BC, Fauser BC, Mol BW, Broekmans FJ. AMH and AFC as predictors of excessive response in controlled ovarian hyperstimulation: a meta-analysis. Hum Reprod Update 2011;17:46–54.
18. Anckaert E, Smitz J, Schiettecatte J, Klein B, Arce JC. The value of antimüllerian hormone measurement in the long GnRH agonist protocol: association with ovarian response, dose adjustments, embryo quality and pregnancy. Hum Reprod 2012;27:1829–39.
19. Arce JC, La Marca A, Klein BM, Andersen AN, Fleming R. Antimüllerian hormone in GnRH antagonist cycles: prediction of ovarian response and cumulative treatment outcome in good prognosis patients. Fertil Steril 2013;99: 1644–53.
20. Nyboe Andersen A, Witjes H, Gordon K, Mannaerts B, Xpect investigators. Predictive factors of ovarian response and clinical outcome after IVF/ICSI following a rFSH/GnRH antagonist protocol with or without oral contraceptive pre-treatment. Hum Reprod 2011;26:3413–23.
21. Nyboe Andersen A, Devroey P, Arce JC. Clinical outcome following stimulation with highly purified hMG or recombinant FSH in patients undergoing IVF: a randomized assessor-blind controlled trial. Hum Reprod 2006;21: 3217–27.
22. Devroey P, Pellicer A, Nyboe Andersen A, Arce JC, Menopur in GnRH Antagonist Cycles with Single Embryo Transfer Trial Group. A randomized assessor-blind trial comparing highly purified hMG and recombinant FSH in a GnRH antagonist cycle with compulsory single-blastocyst transfer. Fertil Steril 2012;97:561–71.
23. Broer SL, van Disseldorp J, Broeze KA, Dolleman M, Opmeer BC, Bossuyt P, et al. Added value of ovarian reserve testing on patient characteristics in the prediction of ovarian response and ongoing pregnancy: an individual patient data approach. Hum Reprod Update 2013;19:26–36.
24. Broer SL, Dólleman M, van Disseldorp J, Broeze KA, Opmeer BC, Bossuyt PM, et al. Prediction of an excessive response in in vitro fertilization from patient characteristics and ovarian reserve tests and comparison in subgroups: an individual patient data meta-analysis. Fertil Steril 2013;100:420–9.e7.
25. Hansen KR, Morris JL, Thyer AC, Soules MR. Reproductive aging and variability in the ovarian antral follicle count: application in the clinical setting. Fertil Steril 2003;80:577–83.
26. Jayaprakasan K, Hopkisson JF, Campbell BK, Clewes J, Johnson IR, Raine-Fenning NJ. Quantification of the effect of pituitary down-regulation on 3D ultrasound predictors of ovarian response. Hum Reprod 2008;23: 1538–44.
27. Sharara FI, Lim J, McClamrock HD. The effect of pituitary desensitization on ovarian volume measurements prior to in-vitro fertilization. Hum Reprod 1999;14:183–5.
28. Yu Ng EH, Chi Wai Chan C, Tang OS, Shu Biu Yeung W, Chung Ho P. Effect of pituitary downregulation on antral follicle count, ovarian volume and stromal blood flow measured by three-dimensional ultrasound with power Doppler prior to ovarian stimulation. Hum Reprod 2004;19: 2811–5.
29. Dewailly D, Lujan ME, Carmina E, Cedars MI, Laven J, Norman RJ, et al. Definition and significance of polycystic ovarian morphology: a task force report from the Androgen Excess and Polycystic Ovary Syndrome Society. Hum Reprod Update 2014;20:334–52.
30. Iliodromiti S, Anderson RA, Nelson SM. Technical and performance characteristics of antimüllerian hormone and antral follicle count as biomarkers of ovarian response. Hum Reprod Update. 2014 Dec 8. pii: dmu062. [Epub ahead of print].
31. Gougeon A. Dynamics of follicular growth in the human: a model from preliminary results. Hum Reprod 1986;1:81–7.
32. Deb S, Jayaprakasan K, Campbell BK, Clewes JS, Johnson IR, Raine-Fenning NJ. Intraobserver and interobserver reliability of automated antral follicle counts made using three-dimensional ultrasound and SonoAVC. Ultrasound Obstet Gynecol 2009;33:477–83.
33. Deb S, Campbell BK, Clewes JS, Raine-Fenning NJ. Quantitative analysis of antral follicle number and size: a comparison of two-dimensional and automated three-dimensional ultrasound techniques. Ultrasound Obstet Gynecol 2010;35:354–60.
34. Broekmans FJ, de Ziegler D, Howles CM, Gougeon A, Trew G, Olivennes F. The antral follicle count: practical recommendations for better standardization. Fertil Steril 2010;94:1044–51.
35. Dewailly D, Andersen CY, Balen A, Broekmans F, Dilaver N, Fanchin R, et al. The physiology and clinical utility of antimüllerian hormone in women. Hum Reprod Update 2014;20:370–85.
36. Gassner D, Jung R. First fully automated immunoassay for antimüllerian hormone. Clin Chem Lab Med 2014;52:1143–52.
37. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 1986;15:413–9.
38. McNamee R. Confounding and confounders. Occup Environ Med 2003;60: 227–34.
39. Weinberg CR. Toward a clearer definition of confounding. Am J Epidemiol 1993;137:1–8.
40. Iliodromiti S, Kelsey TW, Wu O, Anderson RA, Nelson SM. The predictive accuracy of antimüllerian hormone for live birth after assisted conception: a systematic review and meta-analysis of the literature. Hum Reprod Update 2014;20:560–70.

SUPPLEMENTAL FIGURE 1

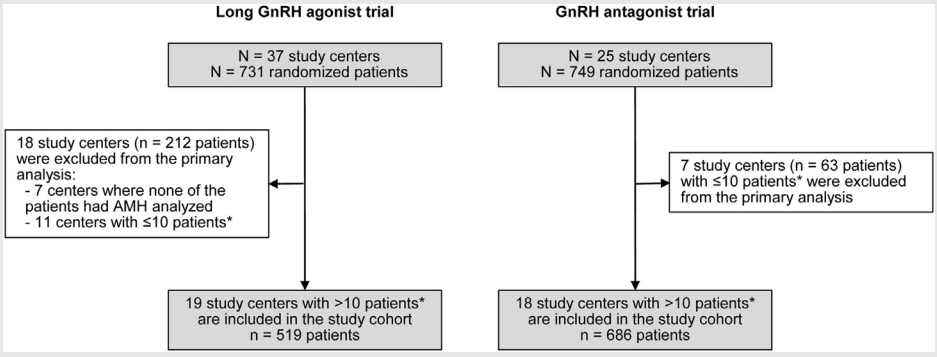


Diagram showing the data selection procedure. *Patients who underwent oocyte retrieval and had assessments of both AMH and AFC at stimulation day 1.

Nelson. AMH vs. AFC as predictor of oocyte yield. Fertil Steril 2015.