

Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level

Dave J. Hendriks, M.D.,^a Ben-Willem J. Mol, Ph.D.,^b László F. J. M. M. Bancsi, Ph.D.,^a Egbert R. te Velde, Ph.D., and ^a Frank J. M. Broekmans, Ph.D.^a

^aDepartment of Reproductive Medicine, Division of Obstetrics, Neonatology and Gynecology, University Medical Center Utrecht, Utrecht; and ^bCentre for Reproductive Medicine, Department of Obstetrics and Gynecology, Academic Medical Center, Amsterdam, The Netherlands

Objective: To assess the predictive performance of the antral follicle count (AFC) as a test for ovarian reserve in IVF patients and to compare this performance with that of basal FSH level.

Design: Meta-analysis.

Setting: Tertiary fertility center.

Patient(s): Patients undergoing IVF.

Intervention(s): None.

Main Outcome Measure(s): Poor ovarian response, nonpregnancy.

Result(s): We identified 11 studies on AFC and an updated total of 32 studies on basal FSH from the literature on the basis of preset criteria. The estimated summary receiver operating characteristic (ROC) curves showed AFC to perform well in the prediction of poor ovarian response. Also, prediction of poor ovarian response seemed to be more accurate with AFC compared with basal FSH. The estimated summary ROC curves for the prediction of nonpregnancy indicated a poor performance for both AFC and basal FSH.

Conclusion(s): Transvaginal ultrasonography is an easy-to-perform and noninvasive method that provides essential predictive information on ovarian responsiveness. The predictive performance of AFC toward poor response is significantly better than that of basal FSH. Therefore, AFC might be considered the test of first choice in the assessment of ovarian reserve prior to IVF. (Fertil Steril® 2005;83:291–301. ©2005 by American Society for Reproductive Medicine.)

Key Words: Antral follicle count, basal FSH, IVF, poor response, pregnancy, meta-analysis

Reproductive aging is associated with a reduction of the primordial follicle pool and loss of oocyte quality (1). The number of follicles leaving the pool of resting follicles to enter the growth phase toward the antral stages of development decreases with increasing age (2).

In an IVF program, ovarian aging is characterized by decreased ovarian responsiveness to gonadotropin administration and lowered pregnancy rates. Correct identification of patients at risk of poor ovarian response by assessment of ovarian reserve before entering an IVF program is important.

Received May 14, 2004; revised and accepted August 13, 2004.

Presented as an oral presentation at the 20th Annual Meeting of the European Society of Human Reproduction and Embryology in Berlin, Germany, June 27–30, 2004.

Reprint requests: D. J. Hendriks, M.D., University Medical Center Utrecht, Department of Reproductive Medicine, Division of Obstetrics, Neonatology and Gynecology, Heidelberglaan 100, Utrecht 3584 CX, The Netherlands (FAX: 31-30-2505433; E-mail: d.hendriks@azu.nl).

It can help physicians to tailor their advice to individual couples and help patients to decide whether to proceed with a costly and often demanding and disappointing IVF treatment.

A recent meta-analysis on the performance of basal FSH level in the prediction of poor ovarian response and failure to become pregnant after IVF showed that a possible clinical application of basal FSH refers to only a minority of patients with extremely high basal FSH levels (3). However, to date basal FSH is still widely used in many fertility centers.

Several other endocrine ovarian reserve tests have been proposed as predictors for IVF outcome in the past decade (4). In the field of quantitative ultrasonography, achievements in the development of tests that assess ovarian reserve have also been reported. After the initial reports by Reuss et al. (5) and Scheffer et al. (6) that transvaginal ultrasonography could detect age-related decreases in follicle counts,

Tomas et al. (7) and Chang et al. (8) introduced the antral follicle count (AFC) as an easy-to-perform and noninvasive method to provide essential information on ovarian responsiveness before initiation of gonadotropin stimulation in IVF.

A low number of small antral follicles is associated with decreased ovarian response during controlled ovarian hyperstimulation for IVF, supporting the concept of reduced numbers of primordial follicles delivering a small antral follicle cohort. Moreover, Chang et al. (8) reported a trend toward lower pregnancy rates in women with few antral follicles. However, this trend lacked statistical significance.

The predictive capacity of AFC for the outcome in IVF patients has not been assessed systematically. There are a limited number of studies that assessed the performance of AFC in predicting the occurrence of nonpregnancy and/or poor response. Unfortunately, most of these studies report on a small number of patients. To overcome this shortcoming, we performed a meta-analysis on the subject.

The aim of the meta-analysis was to determine the predictive capacity of AFC measurements in the prediction of IVF outcome with regard to both ovarian response and pregnancy. Furthermore, we aimed to compare the predictive capacity of AFC measurements resulting from this meta-analysis with the predictive capacity as assessed for basal FSH in an update of a recent meta-analysis (3).

MATERIALS AND METHODS

Search Strategy

We performed a computerized MEDLINE search to identify all studies published up to October 2003 in which AFC was used as a prognosticator for IVF outcome. The search covered the period from January 1996 onward because the first report on AFC was published in 1996 (5). Keywords used were “antral follicle count” or “antral follicle number” and “in vitro fertilization” or “in vitro fertilisation” or “assisted” or “intracytoplasmic” or “intra-cytoplasmatic.”

One investigator (D.H.) read all abstracts of the articles that were identified by the search. To be included, studies had to contain data on AFC and IVF outcome. Data on IVF outcome included poor ovarian response and/or pregnancy. There were no uniform criteria for the definition of poor ovarian response. In this analysis, poor response encompassed cycle cancellation as well as insufficient follicular growth or oocyte yields according to the standards of each study. With respect to pregnancy, data on clinical and ongoing pregnancies were not analyzed separately. In addition, cross-references of the selected studies were checked for other articles meeting the inclusion criteria, and, if they were applicable, these studies were added to the analysis.

Two-by-two tables comparing results of AFC and the occurrence of poor ovarian response and/or pregnancy were constructed independently by two of the authors (D.H. and

F.B.) and in the event of disagreement, the judgment of a third author (B.W.M.) was decisive. The authors of studies in which it was not possible to construct 2×2 tables were asked to supply us with data needed for the construction of 2×2 tables, otherwise these studies were excluded.

The following quality characteristics of each study were registered: [1] sampling (consecutive vs. other), [2] data collection (prospective vs. retrospective), [3] study design (cohort study vs. case-control study), [4] blinding (present or absent), [5] selection bias, and [6] verification bias (9). The definition of poor response or pregnancy was documented, as well as whether an included study reported on only one cycle per couple or on multiple cycles.

For the comparison of AFC and basal FSH, we used an updated version of the recently published meta-analysis on the performance of basal FSH (3). To do so, we extended a previous MEDLINE search for studies reporting on basal FSH in the prediction of poor ovarian response and nonpregnancy in IVF patients. This extended search covered the period from December 1999 to October 2003. If new studies were eligible for meta-analysis, according to the procedures described in the recently published meta-analysis, they were added to the studies that already had been detected.

Analysis

The analysis of the data extracted from the studies was conducted according to a methodology that has been described in detail elsewhere (3, 10). Briefly, for each study sensitivity and specificity were calculated from the published data or on the basis of information provided by the authors. If it was possible to extract data on the performance of AFC at more than one cut-off value, data on all possible cut-off levels were used. To secure the correct weight of studies, the number of patients was divided by the number of cut-off points. Subsequently, sensitivity-specificity points for each study were plotted in a receiver operating characteristic (ROC) space.

Homogeneity of the studies was tested by means of the χ^2 test statistic (11). A summary point estimate of sensitivity and specificity was calculated if homogeneity could not be rejected. If homogeneity was rejected, logistic regression was used to evaluate whether the study characteristics (i.e., consecutive vs. not consecutive; prospective vs. retrospective; cohort study vs. case-control study; blinding, selection bias, and verification bias present vs. absent) influenced the discriminative capacity of the AFC. A *P* value of $<.05$ was considered statistically significant. If one of the study characteristics was found to have a statistically significant impact on the performance of the test, further analysis was performed in subgroups of patients. Analogous to the AFC studies, the same procedure was followed to assess the predictive performance of basal FSH from the updated meta-analysis of Bancsi et al. (3).

If, even after subgroup analysis, homogeneity continued to be absent, a Spearman correlation coefficient was calcu-

lated for the association between sensitivity and specificity to explore possible heterogeneity due to a shift in cut-off levels of the AFC. If there was a negative correlation between sensitivity and specificity as defined by a correlation coefficient of -0.5 or less, a summary ROC curve was estimated by linear regression after logistic transformation (10, 12, 13).

To compare the predictive capacity of AFC and basal FSH, the estimated summary ROC curves for both AFC and basal FSH in the prediction of poor ovarian response and nonpregnancy were tested for statistically significant differences with a linear regression model after logistic transformation of sensitivity and specificity values.

RESULTS

Search Strategy

For AFC, the computerized MEDLINE search detected 35 articles, of which 18 were excluded on the basis of the abstract because they did not report on the capacity of AFC to predict poor ovarian response and/or pregnancy after IVF. From the cross-references of the remaining 17 articles, another 8 studies were identified for further reading (7, 14–20). Thus, 25 studies were selected, of which 17 had to be excluded for following reasons. From 15 studies, 2×2 tables could not be derived (7, 14, 19–31); 1 study reported on AFC during gonadotropin stimulation (32); and 1 study was a review article (4).

We asked the authors of the studies from which we were not able to construct 2×2 tables to provide us with relevant data for the construction of these tables. Three authors responded and provided reliable data (14, 20, 26). Therefore, a total of 11 studies became available for analysis (8, 14–18, 20, 26, 33–35). Of these studies, 2 reported solely on the prediction of pregnancy (14, 35), 1 reported solely on the prediction of poor response (26), and 8 reported on the prediction of both outcomes (8, 15–18, 20, 33, 34). There was no discordance between the two authors who judged the selected articles.

Study characteristics of the included studies are listed in Table 1. Sampling of patients was consecutive in 9 studies, and data collection was prospective in 10 studies. All studies were designed as cohort studies. Only 1 study was blinded (34). Selection bias was present in almost half of all studies (17, 26, 35). None of the detected studies suffered from apparent verification bias. Seven studies reported on the outcome of one cycle per couple, whereas 3 reported on the outcome of multiple cycles. In 8 of 10 studies, only data on clinical pregnancy were available. Clinical and ongoing pregnancies were usually defined as the presence of a viable fetus on ultrasound examination at 6–7 weeks' gestation and 10–12 weeks' gestation, respectively. The definitions of poor response were diverse.

For basal FSH, the updated computerized MEDLINE search detected another 11 studies eligible for the meta-

analysis (16, 36–45). Three studies reported solely on the prediction of poor response (39, 40, 43), 5 solely on the prediction of pregnancy (36, 37, 41, 42, 45), and 3 reported on the prediction of both outcomes (16, 38, 44).

Analysis

Sensitivities and specificities for the prediction of poor ovarian response, as calculated from each study reporting on AFC, are summarized in Table 2. The sensitivity varied between 9% and 95%, whereas the specificity varied between 40% and 97%. Homogeneity for both sensitivity and specificity had to be rejected (both P values $<.001$). For this reason, the calculation of a summary point estimate for sensitivity and specificity was not meaningful.

Logistic regression analysis showed that none of the study characteristics recorded had a statistically significant impact on the reported predictive performance of AFC. For example, whether the design of the study was retrospective or prospective did not influence the prognostic capacity of AFC as estimated by the studies. A plot of sensitivity–specificity points (solid circles) in ROC space is shown in Figure 1. The Spearman correlation coefficient for sensitivity and specificity was -0.57 , which was judged to be sufficient to estimate a summary ROC curve (Fig. 1, dotted line).

For the prediction of nonpregnancy, the sensitivities and specificities of each study are summarized in Table 3. As for ovarian response, homogeneity for sensitivity and specificity had to be rejected, and logistic regression analysis did not indicate a significant association between study characteristics and the performance of AFC. The sensitivity varied between 7% and 60%, whereas the specificity varied between 33% and 98%. A plot of sensitivity–specificity points (solid circles) in ROC space is shown in Figure 2. The Spearman correlation coefficient for sensitivity and specificity was -0.66 , which was judged to be sufficient to estimate a summary ROC curve (Fig. 2, dotted line).

The 11 selected studies on basal FSH from the updated MEDLINE search were added to the selected studies ($n = 21$) of the already published meta-analysis on basal FSH (3). After renewed analysis, sensitivity and specificity points (open circles) for basal FSH in the prediction of poor response and nonpregnancy were added to Figures 1 and 2, respectively. The Spearman correlation coefficients for sensitivity and specificity were -0.77 and -0.88 for poor ovarian response and nonpregnancy, respectively, which were judged to be sufficient to estimate a summary ROC curve (solid lines, Figs. 1 and 2).

Comparison of the two summary ROC curves indicated a better performance of AFC than of basal FSH in the prediction of poor ovarian response ($P = .009$). Although the predictive capacity for the prediction of nonpregnancy overall was very low for both tests, the performance of AFC seemed slightly better than that of basal FSH. This difference, however, was not statistically significant ($P = .55$).

TABLE 1**Characteristics of included studies on AFC.**

Author (reference)	Consecutive	One cycle per couple	Data per	Definition		Diameter follicles (mm)	Ultrasonograph
				Poor response/cancel	Pregnancy		
Bancsi et al. (33)	Yes	Yes	Cycle	<4 oocytes or <3 follicles 18 mm	Clinical/ongoing	2–5	Toshiba Capasee SSA-220A: 7.5-MHz Probe
Yong et al. (20)	No	Yes	Cycle	<4 oocytes or cancel	Clinical	2–10	Toshiba Eccocee: 7-MHz probe
Kupesic et al. (35)	Yes	Yes	Cycle	Not stated	Clinical	Not stated	Combison 530D: 7.5-MHz probe: Kretztech
Ng et al. (17)	Yes	Yes	Cycle	<3 follicles 15 mm	Clinical	Not stated	Aloka SSD-620: 5-MHz probe
Chang et al. (8)	Yes	No	Cycle	<2 follicles 18 mm	Ongoing	2–5	Accuson 120XP/10: 7-MHz probe
Nahum et al. (16)	Yes	No	Cycle	<3 follicles 18 mm	Clinical	2–6	General electric RT-X200: 6.5-MHz probes
Frattarelli et al. (26)	No	Yes	Cycle	<3 follicles	Not stated	2–10	Acuson 128: 7-MHz probe
Frattarelli et al. (34)	Yes	Yes	Cycle	<3 follicles	Clinical	2–10	Acuson 128: 7-MHz probe
Hsieh et al. (15)	Yes	No	Cycle	No oocytes or poor follicular growth	Clinical	2–10	Acuson Aspen: 4-MHz probe
Fisch and Sher (14)	Yes	Yes	Cycle	Not stated	Clinical	Not stated	Not stated
Sharara and McClamrock (18)	Yes	No	Cycle	Not stated	Clinical	2–8	Not stated

Hendriks. Meta-analysis of AFC. Fertil Steril 2005.

TABLE 2

Performance of AFC in the prediction of poor response in IVF patients and shift from pretest to posttest probability of poor response for patients with an abnormal (lower than the cut-off) AFC result

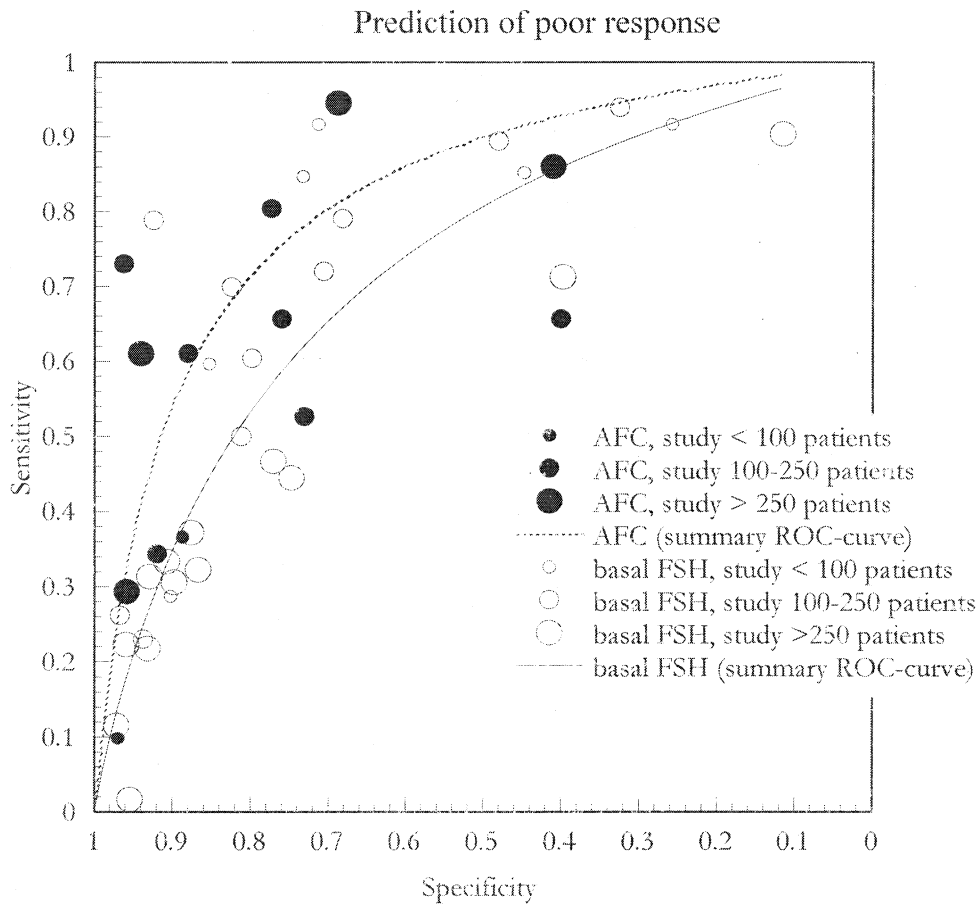
Author (reference)	Cycles (n)	AFC cut-off value (n)	Prediction of poor response			Pre-AFC probability (%)	Post-AFC probability (%)	Proportion of patients/cycles with abnormal AFC (%)
			Sensitivity	Specificity	LR+			
Bancsi et al. (33)	120	≤4	0.61	0.88	5.1	30	69	27
		≤6	0.81	0.77	3.6	30	60	40
		≤8	0.81	0.77	3.6	30	60	40
Ng et al. (17)	128	≤4	0.33	0.92	4.2	2	9	9
		≤6	0.80	0.76	3.3	2	11	27
		≤9	0.80	0.40	1.3	2	5	61
Chang et al. (8)	149	≤3	0.73	0.96	19.7	10	69	11
Shahara et al. (18)	127	≤4	0.53	0.73	1.9	15	26	31
Frattarelli et al. (26)	278	≤10	0.87	0.41	1.5	8	12	61
Hsieh et al. (15)	372	≤3	0.61	0.94	10.0	5	34	9
Frattarelli et al. (34)	267	≤4	0.30	0.96	7.4	9	41	6
Nahum et al. (16)	272	≤6	0.95	0.69	3.1	14	33	39
Yong et al. (20)	47	≤4	0.09	0.97	3.3	23	50	4
		≤6	0.36	0.89	3.3	23	50	17

Note: If a study reported on multiple cut-off values, data for all cut-off values are shown. LR+ = likelihood ratio for a positive test result.

Hendriks. Meta-analysis of AFC. *Fertil Steril* 2005.

FIGURE 1

Summary receiver operating characteristic curves of AFC (dotted line) and basal FSH levels (solid line) in their capacity to predict the occurrence of poor ovarian response in IVF. Solid circles represent the performance of AFC; open circles represent the performance of basal FSH.



Hendriks. Meta-analysis of AFC. *Fertil Steril* 2005.

DISCUSSION

This meta-analysis summarizes the available evidence on the accuracy of AFC in the prediction of poor ovarian response and nonpregnancy in IVF. The performance of AFC for predicting poor ovarian response is good, whereas the performance for predicting nonpregnancy is clearly poor. The literature on AFC as a screening tool for poor IVF outcome has consistently suggested that this test is a useful predictor of ovarian response to hyperstimulation medications. Moreover, it has been suggested that AFC is the best single predictor of ovarian response in assisted reproductive technology cycles (17, 33). The systematic analysis in the present study substantiates these notions.

However, an important matter in meta-analysis is heterogeneity of the included studies, both with respect to aspects of study design as well as clinical characteristics of the couples included. We scored the included studies on relevant

aspects of study design and clinical characteristics and found substantial differences between studies (see Table 1). For example, a possible source of heterogeneity might be the variety of definitions of poor response. In most studies, poor response comprises only canceled cycles. Women with very low numbers of follicles after gonadotropin stimulation whose cycles were not canceled but who also might bear the burden of diminished ovarian reserve were not defined as poor responders (3). However, none of these methodologic parameters had a statistically significant impact on the performance of the tests across the studies.

Judgment on the applicability of AFC in IVF patients should be done in the context of the a priori prospects for IVF outcome, of which female age is the most important predictor (46). The present meta-analysis assessed the performance of AFC in a univariate context, independent of female age. Thus, clinical studies in which the performance

TABLE 3

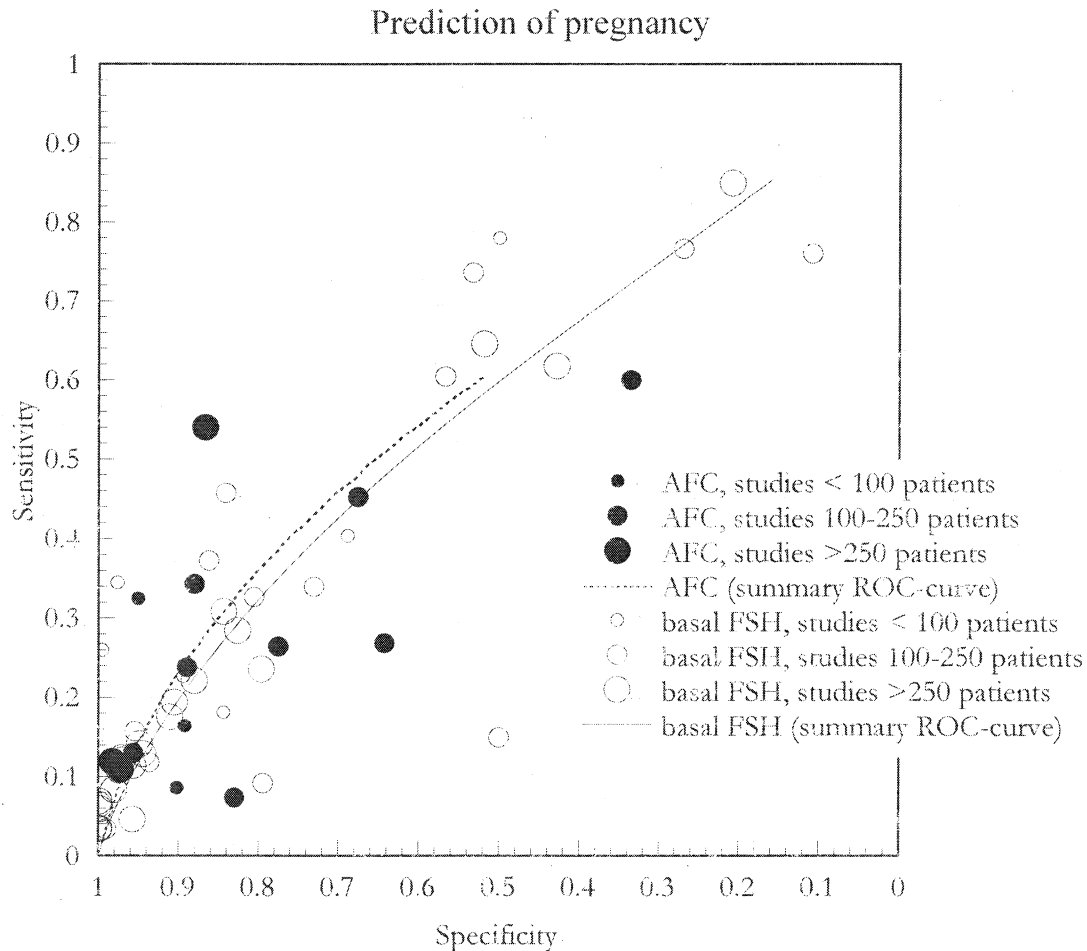
Performance of AFC in the prediction of nonpregnancy in IVF patients and shift from pretest to posttest probability of nonpregnancy for patients with an abnormal (lower than the cut-off) AFC result.

Author (reference)	Cycles (n)	AFC cut-off value (n)	Prediction of nonpregnancy			Pre-AFC probability (%)	Post-AFC probability (%)	Proportion of patients/cycles with abnormal AFC (%)
			Sensitivity	Specificity	LR+			
Bancsi et al. (33)	107	≤4	0.34	0.88	2.9	68	86	27
		≤6	0.45	0.68	1.4	68	75	41
		≤8	0.56	0.56	1.1	68	75	41
Kupesic et al. (35)	56	≤4	0.33	0.96	8.3	61	92	22
		≤6	0.33	0.96	8.3	61	92	22
		≤8	0.33	0.96	8.3	61	92	22
Ng et al. (17)	128	≤4	0.07	0.83	0.4	86	73	9
		≤6	0.26	0.78	1.2	86	88	26
		≤9	0.60	0.33	0.9	86	61	85
Chang et al. (8)	149	≤3	0.13	0.96	3.6	83	94	11
Shahara et al. (18)	127	≤4	0.27	0.64	0.8	56	49	31
Frattarelli et al. (34)	267	≤4	0.11	0.97	4.0	45	76	6
Hsieh et al. (15)	372	≤3	0.12	0.98	6.9	68	94	9
Fisch and Sher (14)	200	≤10	0.24	0.89	2.2	59	76	19
Nahum et al. (16)	272	≤6	0.54	0.87	4.0	64	88	39
Yong et al. (20)	47	≤4	0.08	0.92	0.9	76	75	9
		≤6	0.16	0.90	1.6	79	86	15

Note: If a study reported on multiple cut-off values, data for all cut-off values are shown. LR+ = likelihood ratio for a positive test result.

Hendriks. Meta-analysis of AFC. *Fertil Steril* 2005.

Summary receiver operating characteristic (ROC) curves of AFC (dotted line) and basal FSH levels (solid line) in their capacity to predict the occurrence of nonpregnancy in IVF. Solid circles represent the performance of the AFC; open circles represent the performance of basal FSH.



Hendriks. Meta-analysis of AFC. *Fertil Steril* 2005.

of AFC is compared in a multivariable analysis taking into account its interaction with female age are needed before the true applicability of the AFC can be established. At least in one study, AFC was shown to be the most powerful single predictor of ovarian response, although the addition of other parameters, such as basal FSH and inhibin B, did improve the predictive abilities (33).

The present study compares the accuracy of AFC with that of an updated meta-analysis on basal FSH (3). The performance of AFC for predicting poor ovarian response seemed to be significantly better than that of basal FSH, whereas the performance for predicting nonpregnancy was poor for both AFC and basal FSH.

Assessment of basal FSH, introduced by Scott et al. (47), is generally accepted as a test to assess ovarian reserve (48). However, basal FSH concentrations might vary from cycle

to cycle (49, 50), and the reported threshold values can range from 10 IU/L to 25 IU/L (3, 51). This might lead to variable results in the prediction of ovarian response. Also, the recent meta-analysis on the performance of basal FSH as a screening test to predict ovarian response or nonpregnancy in IVF showed the limited clinical value of this test, especially in the prediction of nonpregnancy (3). A possible clinical application of basal FSH refers to only a minority of patients with extremely high basal FSH levels.

In a recent study by van Rooij et al. (52), it was found that younger patients (≤ 41 years of age) with elevated basal FSH levels have quite acceptable ongoing pregnancy rates, despite a considerable probability of cycle cancellation due to poor ovarian response. Patients with a high basal FSH might include women with high endogenous FSH levels, which are physiologic and not related to ovarian reserve. In young,

normal ovulatory women, for example, a wide range of FSH levels has been demonstrated (53). Additionally, in mothers of familial dizygotic twins, higher FSH levels are associated with an increase in secretory drive of FSH instead of limitations in hormonal feedback, which is the case with reproductive aging (54).

Finally, ovarian response in IVF is clearly dependent on the genotype of the FSH receptor (55, 56). Different variants in receptor genotype have been related to different basal FSH levels and to different numbers of FSH ampoules needed to achieve adequate ovarian response in IVF. We should therefore question the value of routine screening for elevated basal FSH levels without differentiating among the various possible causes.

Because AFC is a rather new ovarian reserve test, and most studies reported that AFC is one of the better predictors of response to ovarian stimulation with exogenous gonadotropins, significant questions remain regarding factors that might influence AFC. As already stated, young patients with a low AFC might be considered as having diminished ovarian reserve, but oocyte quality might still be very acceptable and therefore young patients should not be restricted from treatment (52). Recently, Hansen et al. (57) concluded that an AFC can be reliably performed before or after pituitary downregulation. Although there are moderate differences in AFC between observers, AFC proved to be a useful predictor of stimulation outcome in IVF. The intercycle variability in infertility patients undergoing IVF was greater than the interobserver variability, suggesting that there are biological variations in AFC from cycle to cycle. However, attempts to identify an optimal cycle for stimulation based on basal AFC will not result in significantly better outcomes. This is completely in line with the findings from the study by Bancsi et al. (58), in which it was shown that repeated AFCs did not improve the predictive power of the test.

The finding that AFC performed better than basal FSH seems not to be surprising, because basal FSH levels are influenced in many ways and are potentially susceptible to large variations. Therefore, basal FSH can only act as an indirect measure for the actual cohort size. Antral follicle count should instead be considered a more direct reflection of ovarian reserve and is found to be stable in terms of intercycle variability (59). We are aware of the fact that we used different populations in the comparison of both tests in the present study. This obviously could hamper comparability of the performance of these tests. It would be possible to overcome this problem by limiting the analysis to studies reporting on the predictive capacity of both tests in the same groups of women. However, only two studies satisfied this criterion (16, 33). In both studies, it was concluded that AFC was a better predictor of IVF outcome than was basal FSH.

The potential clinical value of a test has to be considered from the perspective of both false-positive and false-

negative test results. The clinical consequences of an adequate prediction of poor response should be considered before a decision to assess the ovarian reserve with medical tests. A high chance of poor response might lead to the decision to start stimulation with a relatively high dose of gonadotropins. By doing so, it is evident that in the case of an observed poor response there is no question as to whether the dose was sufficiently high. On the other hand, a recent randomized study has suggested that the additional value of a high starting dose might have no benefit in cases of an expected poor response (unpublished observations). For patients with false-positive results, this might implicate an increased risk of exaggerated ovarian response. Likewise, a suboptimal stimulation procedure with the standard regimen might result from a false-negative test. A positive test result in the prediction of nonpregnancy might imply negative counseling or denial of entry into an IVF program. In this case, the impact of a false-positive test result is clearly more dramatic. A low false-positive rate requires excellent specificity, which can only be obtained at the expense of a decreased sensitivity, and will only refer to very few patients because the cut-off level necessary to reach such specificity will be high.

The data from the meta-analysis for basal FSH as a test for poor response and nonpregnancy showed no combinations of high specificity and high, or even intermediate, sensitivity (3). Data from the current meta-analysis indicate that, when AFC is used as a test, only for the outcome variable poor ovarian response are there more combinations of high specificity and intermediate sensitivity, compared with the use of basal FSH. For the outcome variable nonpregnancy, no such combinations were found either for AFC or for basal FSH.

The poor performance in predicting nonpregnancy for both AFC and basal FSH might be not much of a surprise because these tests merely represent the quantitative aspect of ovarian reserve. The occurrence of pregnancy in IVF is largely dependent on oocyte quality, although other factors, partly unknown, will also play a determining role. Oocyte quality cannot be directly assessed for obvious reasons. However, the ability to predict poor response might be a valuable tool for patient counseling because poor responders can expect a lower probability of pregnancy in consecutive cycles (60).

In conclusion, our study shows that the performance of AFC in the prediction of poor ovarian response in IVF is quite adequate and superior to that of the widely used basal FSH. In view of this superior accuracy of AFC, its stable intercycle variability, its low costs, and the practicability of the test, we believe that this test could be preferred over basal FSH if any test for ovarian reserve assessment is to be carried out before IVF.

REFERENCES

1. te Velde ER, Pearson PL. The variability of female reproductive ageing. *Hum Reprod Update* 2002;8:141–54.
2. Gougeon A, Ecochard R, Thalabard JC. Age-related changes of the population of human ovarian follicles: increase in the disappearance rate of non-growing and early-growing follicles in aging women. *Biol Reprod* 1994;50:653–63.
3. Bancsi LF, Broekmans FJ, Mol BW, Habbema JD, te Velde ER. Performance of basal follicle-stimulating hormone in the prediction of poor ovarian response and failure to become pregnant after in vitro fertilization: a meta-analysis. *Fertil Steril* 2003;79:1091–100.
4. Bukman A, Heineman MJ. Ovarian reserve testing and the use of prognostic models in patients with subfertility. *Hum Reprod Update* 2001;7:581–90.
5. Ruess ML, Kline J, Santos R, Levin B, Timor-Tritsch I. Age and the ovarian follicle pool assessed with transvaginal ultrasonography. *Am J Obstet Gynecol* 1996;174:624–7.
6. Scheffer GJ, Broekmans FJ, Dorland M, Habbema JD, Looman CW, te Velde ER. Antral follicle counts by transvaginal ultrasonography are related to age in women with proven natural fertility. *Fertil Steril* 1999;72:845–51.
7. Tomas C, Nuojua-Huttunen S, Martikainen H. Pretreatment transvaginal ultrasound examination predicts ovarian responsiveness to gonadotropins in in-vitro fertilization. *Hum Reprod* 1997;12:220–3.
8. Chang MY, Chiang CH, Hsieh TT, Soong YK, Hsu KH. Use of the antral follicle count to predict the outcome of assisted reproductive technologies. *Fertil Steril* 1998;69:505–10.
9. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, Bossuyt PM. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.
10. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994;120:667–76.
11. SAS Institute. *SAS/STAT user's guide*, version 6. 4th ed. Cary, NC: SAS Institute, 1989:865–9.
12. Midgette AS, Stukel TA, Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. *Med Decis Making* 1993;13:253–7.
13. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993;12:1293–316.
14. Fisch JD, Sher G. The antral follicle count (AFC) correlates with metaphase II oocytes and ART cycle outcome: an update. *Fertil Steril* 2002;78(3 Suppl 1):S90.
15. Hsieh YY, Chang CC, Tsai HD. Antral follicle counting in predicting the retrieved oocyte number after ovarian hyperstimulation. *J Assist Reprod Genet* 2001;18:320–4.
16. Nahum R, Shifren JL, Chang Y, Leykin L, Isaacson K, Toth TL. Antral follicle assessment as a tool for predicting outcome in IVF—is it a better predictor than age and FSH? *J Assist Reprod Genet* 2001;18:151–5.
17. Ng EH, Tang OS, Ho PC. The significance of the number of antral follicles prior to stimulation in predicting ovarian responses in an IVF programme. *Hum Reprod* 2000;15:1937–42.
18. Sharara FI, McClamrock HD. Antral follicle count and ovarian volume predict IVF outcome. *Fertil Steril* 2000;74(3 Suppl 1):S176.
19. Sylvestre CV, Child T, Pirwany I, Tan SL. Baseline ultrasound and serum hormonal prediction of follicular response to gonadotropin stimulation during IVF. *Fertil Steril* 2002;78 Suppl 1:S6–S7.
20. Yong PY, Baird DT, Thong KJ, McNeilly AS, Anderson RA. Prospective analysis of the relationships between the ovarian follicle cohort and basal FSH concentration, the inhibin response to exogenous FSH and ovarian follicle number at different stages of the normal menstrual cycle and after pituitary down-regulation. *Hum Reprod* 2003;18:35–44.
21. Beckers NG, Macklon NS, Eijkemans MJ, Fauser BC. Women with regular menstrual cycles and a poor response to ovarian hyperstimulation for in vitro fertilization exhibit follicular phase characteristics suggestive of ovarian aging. *Fertil Steril* 2002;78:291–7.
22. Chiang CH, Chang MY, Shiau CS, Hou HC, Hsieh TT, Soong YK. Effect of a sonographically diffusely enlarged uterus without distinct uterine masses on the outcome of in vitro fertilization-embryo transfer. *J Assist Reprod Genet* 1999;16:369–72.
23. Chiang CH, Hsieh TT, Chang MY, Shiau CS, Hou HC, Hsu JJ, et al. Prediction of pregnancy rate of in vitro fertilization and embryo transfer in women aged 40 and over with basal uterine artery pulsatility index. *J Assist Reprod Genet* 2000;17:409–14.
24. Dumesic DA, Damario MA, Session DR, Famuyide A, Lesnick TG, Thornhill AR, et al. Ovarian morphology and serum hormone markers as predictors of ovarian follicle recruitment by gonadotropins for in vitro fertilization. *J Clin Endocrinol Metab* 2001;86:2538–43.
25. Ficiocioglu C, Kutlu T, Demirbasoglu S, Mulayim B. The role of inhibin B as a basal determinant of ovarian reserve. *Gynecol Endocrinol* 2003;17:287–93.
26. Frattarelli JL, Lauria-Costab DF, Miller BT, Bergh PA, Scott RT. Basal antral follicle number and mean ovarian diameter predict cycle cancellation and ovarian responsiveness in assisted reproductive technology cycles. *Fertil Steril* 2000;74:512–7.
27. Jarvela IY, Sladkevicius P, Kelly S, Ojha K, Campbell S, Nargund G. Quantification of ovarian power Doppler signal with three-dimensional ultrasonography to predict response during in vitro fertilization. *Obstet Gynecol* 2003;102:816–22.
28. Loverro G, Nappi L, Mei L, Giacomoantonio L, Carriero C, Tartagni M. Evaluation of functional ovarian reserve in 60 patients. *Reprod Biomed Online* 2003;7:200–4.
29. Pohl M, Hohlagschwandtner M, Obruca A, Poschalko G, Weigert M, Feichtinger W. Number and size of antral follicles as predictive factors in vitro fertilization and embryo transfer. *J Assist Reprod Genet* 2000;17:315–8.
30. Popovic-Todorovic B, Loft A, Lindhard A, Bangsboll S, Andersson AM, Andersen AN. A prospective study of predictive factors of ovarian response in 'standard' IVF/ICSI patients treated with recombinant FSH. A suggestion for a recombinant FSH dosage normogram. *Hum Reprod* 2003;18:781–7.
31. Sharara FI, McClamrock HD. The effect of aging on ovarian volume measurements in infertile women. *Obstet Gynecol* 1999;94:57–60.
32. Huang FJ, Chang SY, Tsai MY, Kung FT, Wu JF, Chang HW. Determination of the efficiency of controlled ovarian hyperstimulation in the gonadotropin-releasing hormone agonist-suppression cycle using the initial follicle count during gonadotropin stimulation. *J Assist Reprod Genet* 2001;18:91–6.
33. Bancsi LF, Broekmans FJ, Eijkemans MJ, de Jong FH, Habbema JD, te Velde ER. Predictors of poor ovarian response in in vitro fertilization: a prospective study comparing basal markers of ovarian reserve. *Fertil Steril* 2002;77:328–36.
34. Frattarelli JL, Levi AJ, Miller BT, Segars JH. A prospective assessment of the predictive value of basal antral follicles in in vitro fertilization cycles. *Fertil Steril* 2003;80:350–5.
35. Kupesic S, Kurjak A, Bjelos D, Vujisic S. Three-dimensional ultrasonographic ovarian measurements and in vitro fertilization outcome are related to age. *Fertil Steril* 2003;79:190–7.
36. Bancsi LF, Huijs AM, den Ouden CT, Broekmans FJ, Looman CW, Blankenstein MA, et al. Basal follicle-stimulating hormone levels are of limited value in predicting ongoing pregnancy rates after in vitro fertilization. *Fertil Steril* 2000;73:552–7.
37. Chae HD, Kim CH, Kang BM, Chang YS. Clinical usefulness of basal FSH as a prognostic factor in patients undergoing intracytoplasmic sperm injection. *J Obstet Gynaecol Res* 2000;26:55–60.
38. Chuang CC, Chen CD, Chao KH, Chen SU, Ho HN, Yang YS. Age is a better predictor of pregnancy potential than basal follicle-stimulating hormone levels in women undergoing in vitro fertilization. *Fertil Steril* 2003;79:63–8.
39. Creus M, Penarrubia J, Fabregues F, Vidal E, Carmona F, Casamitjana R, et al. Day 3 serum inhibin B and FSH and age as predictors of

- assisted reproduction treatment outcome. *Hum Reprod* 2000;15:2341–6.
40. Fabregues F, Balasch J, Creus M, Carmona F, Puerto B, Quinto L, et al. Ovarian reserve test with human menopausal gonadotropin as a predictor of in vitro fertilization outcome. *J Assist Reprod Genet* 2000;17:13–9.
 41. Jinno M, Hoshiai T, Nakamura Y, Teruya K, Tsunoda T. A novel method for assessing assisted female fertility: bioelectric impedance. *J Clin Endocrinol Metab* 2000;85:471–4.
 42. Mikkelsen AL, Andersson AM, Skakkebaek NE, Lindenberg S. Basal concentrations of oestradiol may predict the outcome of in-vitro maturation in regularly menstruating women. *Hum Reprod* 2001;16:862–7.
 43. Penarrubia J, Balasch J, Fabregues F, Carmona F, Casamitjana R, Moreno V, et al. Day 5 inhibin B serum concentrations as predictors of assisted reproductive technology outcome in cycles stimulated with gonadotrophin-releasing hormone agonist-gonadotrophin treatment. *Hum Reprod* 2000;15:1499–504.
 44. van der Stege JG, van der Linden PJ. Useful predictors of ovarian stimulation response in women undergoing in vitro fertilization. *Gynecol Obstet Invest* 2001;52:43–6.
 45. Yanushpolsky EH, Hurwitz S, Tikh E, Racowsky C. Predictive usefulness of cycle day 10 follicle-stimulating hormone level in a clomiphene citrate challenge test for in vitro fertilization outcome in women younger than 40 years of age. *Fertil Steril* 2003;80:111–5.
 46. Templeton A, Morris JK, Parslow W. Factors that affect outcome of in-vitro fertilisation treatment. *Lancet* 1996;348:1402–6.
 47. Scott RT, Toner JP, Muasher SJ, Oehninger S, Robinson S, Rosenwaks Z. Follicle-stimulating hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. *Fertil Steril* 1989;51:651–4.
 48. Barnhart K, Osheroff J. Follicle stimulating hormone as a predictor of fertility. *Curr Opin Obstet Gynecol* 1998;10:227–32.
 49. Scott RT Jr, Hofmann GE, Oehninger S, Muasher SJ. Intercycle variability of day 3 follicle-stimulating hormone levels and its effect on stimulation quality in in vitro fertilization. *Fertil Steril* 1990;54:297–302.
 50. Bancsi LF, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. Predicting poor ovarian response in IVF: use of repeat basal FSH measurement. *J Reprod Med* 2004;49:187–94.
 51. Sharara FI, Scott RT Jr, Seifer DB. The detection of diminished ovarian reserve in infertile women. *Am J Obstet Gynecol* 1998;179:804–12.
 52. van Rooij IA, Bancsi LF, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. Women older than 40 years of age and those with elevated follicle-stimulating hormone levels differ in poor response rate and embryo quality in in vitro fertilization. *Fertil Steril* 2003;79:482–8.
 53. Schipper I, de Jong FH, Fauser BC. Lack of correlation between maximum early follicular phase serum follicle stimulating hormone concentrations and menstrual cycle characteristics in women under the age of 35 years. *Hum Reprod* 1998;13:1442–8.
 54. Lambalk CB, Boomsma DI, De Boer L, De Koning CH, Schoute E, Popp-Snijders C, et al. Increased levels and pulsatility of follicle-stimulating hormone in mothers of hereditary dizygotic twins. *J Clin Endocrinol Metab* 1998;83:481–6.
 55. Perez MM, Gromoll J, Behre HM, Gassner C, Nieschlag E, Simoni M. Ovarian response to follicle-stimulating hormone (FSH) stimulation depends on the FSH receptor genotype. *J Clin Endocrinol Metab* 2000;85:3365–9.
 56. Sudo S, Kudo M, Wada S, Sato O, Hsueh AJ, Fujimoto S. Genetic and functional analyses of polymorphisms in the human FSH receptor gene. *Mol Hum Reprod* 2002;8:893–9.
 57. 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.
 58. Bancsi LF, Broekmans FJ, Looman CW, Habbema JD, te Velde ER. Impact of repeated antral follicle counts on the prediction of poor ovarian response in women undergoing in vitro fertilization. *Fertil Steril* 2004;81:35–41.
 59. Kwee J, Schats R, McDonnell J, Lambalk CB, Schoemaker J. Intercycle variability of ovarian reserve tests: results of a prospective randomized study. *Hum Reprod* 2004;19:590–5.
 60. Klinkert ER, Broekmans FJ, Looman CW, te Velde ER. A poor response in the first in vitro fertilization cycle is not necessarily related to a poor prognosis in subsequent cycles. *Fertil Steril* 2004; 81:1247–53.