Comparing anti-Müllerian hormone (AMH) and follicle-stimulating hormone (FSH) as predictors of ovarian function

We compared predictive values of anti-Müllerian hormone (AMH) and baseline FSH with respect to IVF cycle outcomes based on oocyte numbers retrieved and number of clinical pregnancies established. In 76 IVF cycles investigated, AMH was clearly superior in predicting IVF outcomes in comparison with FSH. (Fertil Steril® 2009;91:1553–5. ©2009 by American Society for Reproductive Medicine.)
Key Words: Ovarian function, ovarian reserve, in vitro fertilization (IVF), anti-Müllerian hormone (AMH), follicle-stimulating hormone (FSH)

Anti-Műllerian hormone (AMH) and FSH have independently been demonstrated to reflect ovarian reserve. How they compare as predictors is not yet established. In this report we compare the utility of AMH and FSH in predicting retrieved oocyte numbers and clinical pregnancy rates during IVF.

Seventy-six consecutive women, for whom AMH and FSH levels were available in retroactive chart review in association with IVF cycles performed at the Center for Human Reproduction, New York, between January 2007 and March 2008, represented the study population. The median age of analyzed patients was 37 years (range, 17–45 years), with most being Caucasian and nulliparous. Approximately half met classical criteria for a diagnosis of diminished ovarian reserve, and approximately one-quarter for so-called unexplained infertility, which is defined by absence of tubal disease, a partner with normal semen analysis, and baseline FSH <12 mIU/mL. The remaining patients had either tubal or male factor infertility. AMH was drawn at convenience, independent of cycle day and within at most 6 weeks from subsequent IVF cycle start. FSH and E_2 levels were obtained in the morning on days 2 or 3 of the cycle. In cases of multiple available results in one cycle, the least favorable result was used in this analysis. Ovulation induction followed a standard protocol of agonist flare (leuprolide acetate 40 μg per day) on day 2 of menses, followed 3 days later by 300–450 IU of gonadotropins (from varying manufacturers), with FSH preponderance and at most 150 IU of hMG per day of stimulation.

Received July 9, 2008; revised August 28, 2008; accepted September 11, 2008; published online February 12, 2008.
D.H.B. has nothing to disclose. A.W. has nothing to disclose. N.G. has nothing to disclose.
This study was supported by the Foundation for Reproductive Medicine, New York.
Reprint requests: David H. Barad, M.D., M.S., Center for Human Reproduction, 21 East 69th Street, New York, New York 10021 (FAX: 212-994-4499; E-mail: dbarad@thechr.com).

Cycle and patient characteristics were extracted from an electronic database, which is only accessible to the clinical investigators. Pregnancy rates were calculated cumulatively for time of treatment and adjusted for number of IVF cycles performed.

Assays of E_2 and FSH were performed using the Automated Chemiluminescence System (ACS: 180, Bayer Health Care, Tarrytown, NY). Serum AMH was obtained through a commercially available assay, which involves an enzymatically amplified two-site immunoassay, DSL-10-14400 active MIS/AMH ELISA (Esoterix Endocrinology, Casabasas Hills, CA).

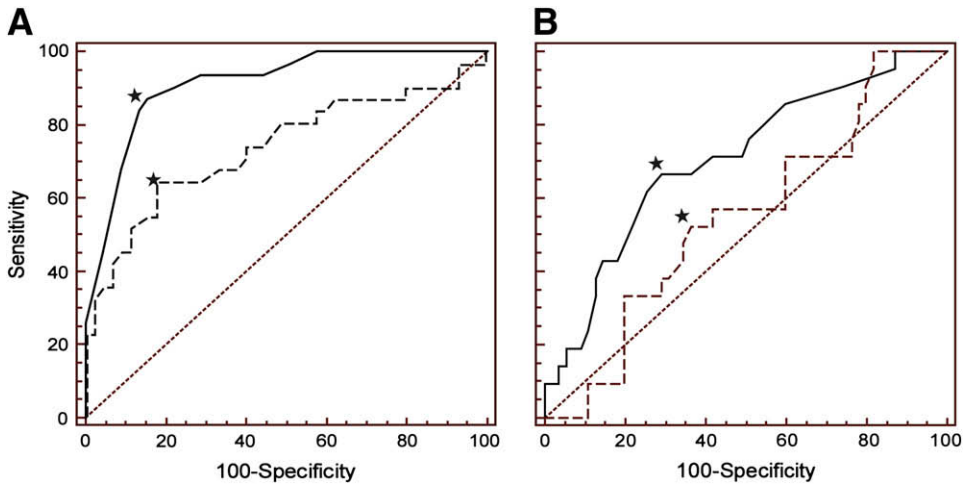
Statistical analysis was performed using SPSS, version 15.0 (Chicago), and MedCalc for Windows, version 9.5.2.0 (MedCalc Software, Mariakerke, Belgium).

Continuous values are presented as mean $\pm$ SD. Variables that did not conform to normality were log converted and back-transformed. They are presented as mean and 95% confidence intervals (CIs) of the mean. Differences between normally distributed variables were tested with analysis of variance or covariance. Differences between groups of variables not conforming to normality were tested for with the Mann-Whitney test, and $P<.05$ was considered statistically significant.

Regression analysis was used to calculate the continuous relationship of retrieved oocytes to serum AMH and baseline FSH. Residual analysis was used to confirm that the regression models conformed to the assumptions of normal distribution and homogenous variance. Receiver operator characteristic curves (ROCs) were then calculated. The conditions tested for were clinical pregnancy or production of fewer than four oocytes. The prevalence of these conditions was estimated from the ratio of cases in the positive and negative groups. The results provide a of list cutoff values, with corresponding sensitivity and specificity of the test, and the positive and negative likelihood ratio. Selected cutoff values were then used in logistic regression models,

FIGURE 1

(**A**) ROCs for AMH and baseline FSH for prediction of retrieval of more than four oocytes. The small stars mark the points of maximal sensitivity and specificity, which are equivalent to 0.5 ng/mL AMH (87% sensitivity and 84% specificity) and 12 mIU/mL baseline FSH (65% sensitivity and 82% specificity. (**B**) ROCs for AMH and baseline FSH for prediction of ongoing clinical pregnancy. The small stars mark the points of maximal sensitivity and specificity, which are equivalent to 1.0 ng/mL AMH (62% sensitivity and 75% specificity) and 11 mIU/mL baseline FSH (62% sensitivity and 40% specificity). The solid line shows AMH, and the dashed line shows FSH.



Barad. AMH and FSH to predict ovarian response. Fertil Steril 2009.

adjusted for age, to calculate age-adjusted odds ratios (ORs) of each outcome for AMH and FSH.

The study underwent expedited Institutional Review Board review, as this study involved only analyses of anonymous medical records. At their initial consultation, all of the center's patients sign a general consent, which allows use of their anonymous medical records for quality-control and research purposes.

Multiple linear regression analysis was used to develop a model for predicting oocyte recovery. Independent variables included in the regression model were age, AMH, FSH, E_2, parity, and number of IVF cycles. In the full model, only AMH demonstrated a significant (partial) effect ($\beta = 4.0 \pm 0.5$; $t = 7.64$; $P<.001$). This multiple regression model, with AMH adjusted for age, FSH, E_2, parity, and IVF cycles, accounted for 57% of the variance in oocyte collection (F [6, 69] = 15.5, $P<.001$; $R^2 = 0.57 \pm 4.07$ [SEM]). Figure 1A compares ROCs for serum AMH (0.90 ± 0.04; 95% CI, 0.81–0.96) and FSH (area under the curve [AUC], 0.73 ± 0.06; 95% CI, 0.61–0.82) for prediction of recovery of fewer than four oocytes. Exactly half (n = 38) of patients had fewer than four oocytes retrieved. As the figure demonstrates, the area under the AMH ROC curve is significantly greater than that of the FSH ROC curve ($z = 2.9$, $P=.004$).

The effect of age on this relationship was investigated by dividing patients into two groups: those under 38 and those 38 or older. Among 37 women in the older group, AMH continued to outperform FSH as a better predictor of fewer than four oocytes retrieved (AMH AUC, 0.89 ± 0.06; FSH AUC, 0.65 ± 0.09; $z = 2.28$, $P=.023$). In the younger group, AUCs for AMH and FSH were not significantly different ($z = 1.63$, $P=.10$).

An AMH cutoff of ≤0.5 ng/mL has a sensitivity of 87% and specificity of 84% for prediction of retrieval of fewer than four oocytes. The positive and negative predictive values for retrieval of fewer than four oocytes at this cutoff are 79.4% and 90%, respectively. An FSH of >12.0 mIU/mL, in contrast, has sensitivity and specificity of 64.5% and 82.2%, respectively, and positive and negative predictive values of 71.4% and 77%.

Using these cutoffs, logistic regression for production of more than four oocytes, adjusted for age, demonstrated that patients with an AMH of >0.5 ng/mL (OR, 32.6; 95% CI, 8.444–126.2, $P<.0001$) and/or FSH ≤ 12 mIU/mL (OR, 5.6; 95% CI, 1.9–16.3; $P=.002$) were more likely to obtain more than four oocytes at retrieval.

Women who achieved pregnancy were younger (34.3 ± 5.3 vs. 37.7 ± 5.3 years; $P<.01$), had higher AMH levels (1.12 ng/mL [0.68, 1.85] vs. 0.52 ng/mL [0.39, 0.69]; $P=.02$], but similar FSH levels (10.4 mIU/mL [8.7, 12.1] vs. 10.9 mIU/mL [9.3, 12.8]; $P=.20$), and produced more oocytes per cycle (8.9 ± 5.8 vs. 5.7 ± 5.8; $P=.04$).

Figure 1B demonstrates ROC curves constructed to assay the comparative predictive values of FSH and AMH based

on 21 patients with ongoing clinical pregnancy and 55 patients who did not achieve pregnancy. The AMH ROC curve (AUC, 0.71 ± 0.98) was significantly better than the FSH ROC curve (AUC, 0.55 ± 0.07) in predicting pregnancy ($z = 2.21$, P=.03). In fact, FSH appeared to have no predictive value for pregnancy (AUC, 0.55±0.07, not significant), while the ROC curve for AMH demonstrates a significant relationship between serum AMH and establishment of a clinical pregnancy (AUC, 0.71 ± 0.07; P=.006). In predicting clinical pregnancy, an AMH level of 1.0 ng/mL has a sensitivity and specificity of 62% and 75%, respectively, and positive and negative predictive values of 48% and 84%. An AMH of greater than 1.0 ng/mL represents 4 times greater odds of achieving clinical pregnancy (OR, 3.9; 95% CI, 1.3–11.9; P=.02).

This study demonstrates that randomly drawn AMH levels are far better predictors of response to ovarian stimulation and of clinical pregnancy than baseline FSH levels. This should not be a complete surprise since FSH is known as a relatively good predictor of response to ovarian stimulation (1) but as a poor predictor of pregnancy (2). Moreover, recent preliminary reports suggested better predictive capabilities for pregnancy for AMH (3).

It has recently been suggested that, as tools to predict ovarian reserve, ROC curves do not yield threshold values of clinical utility (4). This study demonstrates that ROC curves may be used in just such a fashion. They display continuous associations between sensitivity and specificity. If a cutoff point for higher specificity is desired, at the expense of poorer sensitivity, a different cutoff point may be chosen. For instance, an FSH cutoff point of >18 mIU/mL would move to the right of the curve and provide greater specificity (98%) but far less sensitivity (26%). This trade-off between sensitivity and specificity exists no matter what method is used to choose the cutoff point.

Choosing a higher cutoff point allows for greater confidence in recommending against further ovulation induction and in favor of donor eggs. The same cutoff point would, however, have absolutely no utility in attempts to discover incipient transitions toward decreased ovarian reserve, which still may be amenable to ovulation induction.

Since this study group comprised only 76 individuals, the ROC analysis of FSH in predicting clinical pregnancy may not have had adequate statistical power. Since analysis of AMH data was performed on the same data set and did provide useful information concerning oocyte numbers and chance of pregnancy, one can infer that AMH is the more powerful clinical tool. This conclusion is further confirmed by the observation that AMH appears superior, especially among older women, in whom accurate assessments of ovarian reserve would appear even more important. Serum AMH, of course, decreases with increasing age, and only a few women over age 38, therefore, had AMH levels above 1.0 ng/mL.

Adding FSH to AMH, somewhat surprisingly, did not improve results in the predictor models used here. The likely explanations are that AMH and FSH are highly correlated (5) and, as we demonstrated, predictive values of FSH improve when used in an age-specific fashion (6). Age-specific AMH levels may, therefore, be expected to have even superior predictive values, and age-specific FSH in combination with age-specific AMH may end up offering the most reliable way of determining ovarian reserve. Accurate assessment of ovarian function represents one of the big remaining challenges in fertility practice, and, in view of our above reported findings, the addition of AMH as a routine infertility test appears to be warranted.

David H. Barad, M.D., M.S.[a,b,c]
Andrea Weghofer, M.D., Ph.D.[a,d]
Norbert Gleicher, M.D.[a,e]
[a] *Center for Human Reproduction, New York;* [b] *Department of Epidemiology and Social Medicine and* [c] *Department of Obstetrics, Gynecology, and Women's Health, Albert Einstein College of Medicine, Bronx, New York;* [d] *Department of Obstetrics and Gynecology, Allgemeines Krankenhaus of Vienna University School of Medicine, Vienna, Austria; and* [e] *Department of Obstetrics, Gynecology, and Reproductive Sciences, Yale University School of Medicine, New Haven, Connecticut*

REFERENCES

1. Toner JP, Philput CB, Jones GS, Muasher SJ. Basal follicle-stimulating hormone level is a better predictor of in vitro fertilization performance than age. Fertil Steril 1991;55:784–91.
2. 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.
3. Elgindy EA, El-Haieg DO, El-Sebaey A. Anti-Müllerian 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.
4. Scott RT Jr, Elkind-Hirsch KE, Styne-Gross A, Miller KA, Frattarelli JL. The predictive value for in vitro fertility delivery rates is greatly impacted by the method used to select the threshold between normal and elevated basal follicle-stimulating hormone. Fertil Steril 2008;89:868–78.
5. Singer T, Barad DH, Weghofer A, Gleicher N. Correlation of antimüllerian hormone and baseline follicle-stimulating hormone levels. Fertil Steril. Published online 13 June 2008 [Epub ahead of print].
6. Barad DH, Weghofer A, Gleicher N. Age-specific levels for basal follicle-stimulating hormone assessment of ovarian function. Obstet Gynecol 2007;109:1404–10.