dard stimulation protocol, some may benefit from a higher stimulation dose. However, if a poor response is likely, as predicted by a low, antral follicle count, increasing the dose of FSH will probably not prevent a poor or absent response. The chances of pregnancy would thus be low even in younger women with elevated basal FSH levels.

We appreciate the attention drawn by Check to earlier studies affirming the idea that an abnormal basal FSH level in younger women must be considered with caution and should be thought of as a signal and not a diagnosis. Whether these patients will have diminished ovarian reserve must be concluded on the basis of further testing. The diagnosis of diminished ovarian reserve can be made only in case of poor or absent response to maximum ovarian stimulation for IVF.

F. J. Broekmans, Ph.D.
I. A. J van Rooij, M.D.
E. R. te Velde, Ph.D.
Department for Reproductive Medicine
Division of Perinatology and Gynecology
University Medical Centre
Utrecht, The Netherlands
May 2, 2003

doi:10.1016/S0015-0282(03)00962-2

Editorial Commentary

Ovarian reserve—drawing distinctions between the quality and numbers of oocytes

The group from the University Medical Center in Utrecht and Erasmus University in Rotterdam. The Netherlands, have published several good articles on the topic under discussion. In fact, their March 2000 article published in this journal on the prognostic value of serum FSH in predicting ongoing pregnancy rates in IVF is a reference point (1). However, this more recent publication leaves me a little puzzled. From the start, it is not clear why the two different groups—>41 years, normal FSH level vs. <41 years, elevated FSH level—were being compared (2). The authors did not incorporate any element of control or statistical power so that the evidence supporting a relationship can be interpreted. If one is trying to dissect out the interrelationships of age and serum FSH levels in predicting a spectrum of ovarian based responses, the study design serves only to confound. Alternatively, drawing meaningful distinctions between the different ovarian dependent response variables (cycle cancellations and <4 oocytes, pregnancy rates, implantations rates, egg numbers, etc.) would require a much larger sample and corrections for multiple categories of events that are not statistically independent. It appears as if there are too many dependent variables for the numbers of subjects, and several of the outcome events appear to have small numbers (<15 to 20). The results are certainly interesting, but need to be tested in a new study with a larger sample and a hypothesis that controls for age and FSH levels. This is especially important, since elevated gonadotropins and indirectly "abnormal gametes" have been reported as an independent risk factor for aneuploidy, dizygotic twinning, and recurrent spontaneous abortion (3, 4). These reports serve to indicate that drawing distinctions between the numbers and quality of oocytes may be a tricky business.

For interrelationships, a 2 × 2 factorial analysis with adjustment for other potential confounders would be necessary to determine whether serum FSH level is incremental with age in predicting a diminished ovarian response. Using this format, one can analyze each of the two variables separately and determine the presence or absence of interaction between the two in predicting ovarian compromise. The detection of interaction or effect modification is especially important if the prognostic value of one variable (FSH) varies at different levels of the second variable (age). However, one of the potential problems in conducting any type of multiple logistic regression analysis in this setting is that age and serum FSH may be so highly correlated with each other that collinearity occurs. If the two variables (age and serum FSH) are highly correlated (>.80) and contain redundant information, it is difficult to assign the effects of changes in the dependent variables (ovarian function) to one or the other of the predictor variables (age or serum FSH). In some samples with extreme collinearity between the predictor variables, the effect of one variable (age or serum FSH) may cancel out the other or make both of them nonsignificant. The introduction of other predictor variables (basal antral follicle counts, clomiphene citrate challenge test, inhibin B, etc.) into the model may only serve to increase the instability of the regression coefficients when high collinearity is present between age and serum FSH. Under these circumstances, the values of serum FSH and age tend to change together. This makes it difficult to attribute changes in the dependent variable (ovarian function) uniquely to either of the independent variables. This coupling and sharing of information between age and serum FSH about ovarian reserve could be the principal reason that the relative importance of age and serum FSH as predictors of diminished ovarian function varies from study to study.

The higher cancellation rates in the younger group (<41 years) in the study by van Rooij and colleagues may also reflect the asymmetric distribution of the serum FSH values in the younger group where the range includes values from 15.1 IU/L to values as high as 44.0 IU/L. With this positively skewed distribution, the values that exceed the median value (18.1 IU/L) are spread out over a wide range of values. The median as a measure of central tendency is insensitive to the more extreme values. These extreme values in the tail of the distribution are likely to represent patients who may be

"true positives" with clinically significant ovarian compromise.

Whether or not a positive test for FSH provides "convincing evidence" of ovarian compromise ultimately depends on the prior probability of diminished ovarian function in the population under investigation. The authors have applied the test (serum FSH) to a gradient of population subgroups (22.3 to 45.5 years) with differing prevalences of ovarian compromise. The predictive value of the test is not independent of the setting in which the test is used. With this case mix, the positive predictive value of the serum FSH test is going to be much lower in the younger age groups, with more of the normal patients being misclassified as abnormal. Alternatively, when a highly sensitive test is applied to a population in which the disease prevalence is high (>41 years), negative results will be largely false negatives. Because of this differential misclassification, the performance characteristics of the screening test would predict greater pregnancy success rates in the younger group. Similar thoughts about the decrease in specificity in the younger test population are echoed by the authors and correspondent.

In addition, when the disease being detected by the diagnostic test has a low prevalence rate, a prospective design can require an enormous sample of patients in order to obtain an unbiased estimate of test accuracy. For single test studies, the sample size is usually estimated on the desired width of the confidence interval for sensitivity. Differences in the prevalence of ovarian compromise at the extremes of reproductive age may range from <1% to >35% of the respective populations. In a younger population with a prevalence of ovarian compromise approaching 0.5%, it would require approximately 4,000 women to accrue even 20 women with ovarian compromise. Few studies are able to provide this sample size, and misclassification errors are more frequent in younger patients (5). In general, the higher pregnancy rates in younger patients with FSH levels above a given threshold reflect the lower positive predictive value of the test in this population stratum, whereas the increased pretest probability of ovarian compromise in the older age group raises the positive predictive value of the screening test.

In studies of this type, in addition to sample size considerations, it is frequently helpful to stratify the clinical population based on the conjectured prevalence of the disease. Stratification allows a better estimate of test accuracy when accuracy may differ for different patient subpopulations (5). In the final analysis, the value of a test will always depend strongly on the population on which it is used.

As this commentary went to press, an excellent meta-analysis on the performance of serum FSH as a single test in predicting a poor ovarian response was published by the same group from the University Medical Center in Utrecht and Erasmus University in Rotterdam (6). The Utrecht–Rotterdam paper is especially valuable, since it compares the results in a large number of observational studies (n = 21) with samples sizes ranging from <100 patients to studies with >1,000 patients (6). The authors conclude that the criterion to establish a positive diagnosis must be strict, since the cost of a false-positive result is disproportionately high; for example, a false-positive test in a patient with normal ovarian reserve might lead to discontinuation of further attempts at pregnancy. Given the need to avoid "false alarms," the clinical application of the test might apply to only a few patients with extremely high basal FSH levels. The cost of testing the whole population to find these few patients limits the use of FSH testing. It is unfortunate that the system has a fixed capacity to distinguish between positive and negative events. One cannot increase the true-positive proportion without increasing the false-positive proportion.

Paul G. McDonough, M.D.
Associate Editor, Letters
Department of Obstetrics and Gynecology
Medical College of Georgia
Augusta, Georgia
May 16, 2003

References

1. Bansci 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.
2. van Rooij IA, Bansci 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.
3. Freeman SB, Yang Q, Allran K, Taft LF, Sherman SL. Women with a reduced ovarian complement may have an increased risk for a child with Down's syndrome. Am J Hum Genet 2000;66:1680–3.
4. Kline J, Kinney A, Levin B, Warburton D. Trisomic pregnancy and earlier age at menopause. Am J Hum Genet 2000;67:395–404.
5. Obuchowski NA, Zhou XH. Prospective studies of diagnostic test accuracy when disease prevalence is low. Biostatistics 2002;3:477–92.
6. Bansci 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.

doi:10.1016/S0015-0282(03)00963-4

Sperm morphology and rate of blastomere cleavage: correlation?

To the Editor:

The article by Salumets et al. (1) presents an alternative approach to assessing the quality of spermatozoa and oocytes, and their influence on embryonic development. Briefly, oocytes retrieved from donors participating in an oocyte donation program were divided between two recipient couples each time, and the oocytes were inseminated with sperm from the male partner of those recipient couples.

The authors concluded that oocyte quality influences embryo morphology, whereas spermatozoa and oocyte quality