analysis. The design of our study was not to compare the known markers of ovarian reserve to ART outcome. However, we did evaluate the correlation between the ultrasound findings of ovarian diameter and antral follicle count with known factors of ovarian reserve. Our data demonstrated that both ovarian parameters correlated directly with the number of recovered oocytes, basal estradiol levels, and peak estradiol levels, and inversely with ampules of gonadotropins, days of stimulation, patient age, day 3 FSH levels, and FSH:LH ratio. Therefore, these ultrasound parameters represent additional measures of ovarian reserve.

We would like to thank Dr. Ng for disclosing an error in the units for basal estradiol levels in Tables 2 and 3 of our published article. We agree these units should read pg/mL, not mIU/mL. Again, we would like to thank Dr. Ng for his interest in our article. We look forward to publishing more articles on this topic.

John L Fattarelli, M.D.
Department of Obstetrics and Gynecology
Tripler Army Medical Center
Honolulu, Hawaii
November 25, 2000

References

1. Frattarelli JL, Lauri-Costa DF, Miller BT, Bergh PA, Scott RT Jr. 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.
2. Scheffer GJ, Broekmans FJM, Dorland M, Habbema JDF, Looman CWN, teVelde ER. Antral follicle counts by transvaginal ultrasonography are related to age in women with proven natural fertility. Fertil Steril 1999;72:845–51.
3. Toner JP, Philput CB, Jones GS, Muasher SJ. Basal follicle stimulation hormone level is a better predictor of in vitro fertilization performance than age. Fertil Steril 1991;55:784–91.
4. Scott RT, Toner JF, Muasher SJ, Oehninger SC, Robinson S, Rosenwaks Z. Follicle stimulation hormone levels on cycle day 3 are predictive of in vitro fertilization outcome. Fertil Steril 1989;51:651–4.

PII S0015-0282(01)01758-7

Selection of controls—always a challenge

To the Editor:

The selection and interpretation of screening tests contribute to optimal patient care. The article by van Montfrans et al. (1) has raised several issues that deserve comment on how a study about a screening test should be done.

In Dr. van Montfrans' study, they accepted an arbitrary cutoff value for FSH without analyzing the values with a receiver-operating characteristic (ROC) curve. Their reasoning was that this cutoff value identified low responders during IVF in their program. If an optimal cutoff value is not used, the results may be misleading, which we believe happened in this study. The pregnancy rates were reported not to differ between the group with elevated FSH concentrations and the controls. How can we be sure that the results would not have been different if a cutoff value of 12.0, 15.0, or 20.0 had been chosen?

There are two types of studies to evaluate the value of a screening test: case-control studies and prevalence studies (2). In this case-control study, the authors have first selected an arbitrary cutoff value to define normal and abnormal results for their screening test and, secondly, they have evaluated the predictive value of basal FSH screening in their patient population based on this assumption. However, in case-control studies, the starting point should be the outcome of interest, i.e., ovarian response to stimulation or pregnancy rates, both of which indirectly reflect ovarian reserve (2). Thereafter, the predisposing characteristics should be retrospectively looked at. With this type of design, significance of basal FSH concentrations can be analyzed. However, a cutoff value usually can not be determined. Additionally, we believe that their control group had been selected inappropriately, even for a case-control study. Their elevated FSH population ($n = 50$) had been enrolled from January 1, 1995 through March 31, 1997. However, they included only the first 50 age-matched patients (out of a total population of $n = 2,009$) as controls. Considering the difference in success rates in ART over a 26-month period, the choice of each control should have also been "time-matched," at least to the month of enrollment of the patient with an elevated FSH value.

If the goal is to establish a cutoff value for a general subfertility population, the authors should have included the entire subfertile population, which fulfilled all inclusion and exclusion criteria, instead of choosing 50 age-matched controls. Thereafter, if a graph were constructed, where the vertical axis denotes sensitivity for different cutoffs and the horizontal axis displays one-minus-specificity for the same cutoff values and the points in such a graph were connected, a ROC curve could be drawn (2). This curve could also be drawn to assess the predictive value of the patient's age. In this manner, the authors would not only be able to analyze the value of basal FSH concentrations as a screening method for decreased ovarian reserve, but they would also be able to compare the value of different tests, such as basal FSH concentrations and age, by examining the area under each curve. If the tests were found to be predictive, they would also be able to find an optimal cutoff value by the use of this curve (2). Contrary to the authors' method, optimal cutoff values should be determined by each center after evaluation of the distribution of their own data, which has also been mentioned in one of their references (Ref. 1) (3). However, when a large clinical volume is not available, the authors should first correlate their assay to the Leeco Diagnostics assay, which was originally used to determine the optimal cutoff value (3, 4), then they would be able to determine their corresponding cutoff level, predictive of decreased ovarian reserve.

The authors mentioned that the chosen cutoff value identifies low responders during IVF in their program. We wonder whether they may have used different protocols for improving ovarian response in these "poor responders" undergoing IVF. If this is the case, it may have accounted for the lack of significant difference in their pregnancy rates.

Bert Scoccia, M.D.
Koray Elter, M.D.
Linda R. Nelson, M.D., Ph.D.
Division of Reproductive Endocrinology and Infertility
Department of Obstetrics and Gynecology
University of Illinois College of Medicine
Chicago, Illinois
October 26, 2000

References

1. van Montfrans JM, Hoek A, van Hooff MHA, de Koning CH, Tonch N, Lambalk CB. Predictive value of basal follicle-stimulating hormone concentrations in a general subfertility population. Fertil Steril 2000;74: 97–103.
2. Boardman LA, Peipert JF. Screening and diagnostic tests. Clin Obstet Gynecol 1998;41:267–74.
3. Scott RT Jr, Hofmann GE. Prognostic assessment of ovarian reserve. Fertil Steril 1995;63:1–11.
4. Hofmann GE, Sosnowski J, Scott RT, Thie J. Efficacy of selection criteria for ovarian reserve screening using the clomiphene citrate challenge test in tertiary fertility center population. Fertil Steril 1996;66:49–53.

PII S0015-0282(01)01759-9

Reply of the Authors:

Dr. Scoccia and colleagues have made several remarks in relation to our findings that need further clarification.

Our study (1) was basically a follow-up study in patients with signs of decreased ovarian reserve. We used a *nested* case-control design, which differs importantly from a regular case-control study. Indeed, in a regular case-control study, the starting point is the outcome of interest (in this case, pregnancy rates), and the predisposing characteristics can be retrospectively identified. In a *nested* case-control study however, follow-up is evaluated by conducting a case-control study within a cohort of patients, rather than including the entire cohort population. The starting point in a nested case-control study is exposure (in this case elevated basal FSH values), as it is in a follow-up study (2).

The nested case-control design was chosen for several reasons. First, this study design is time- and cost-efficient because it does not evaluate the entire population with negative test results, whereas it does evaluate the whole patient group with positive test results. A nested case-control study produces the same findings as regular follow-up studies with nearly the same level of precision (2). Second, the goal of this study was to evaluate a previously identified cutoff value for poor ovarian response in our IVF population (identifying the highest 2.5% of basal FSH values) in a general subfertility population. It is not likely that our results would have been different if a higher cutoff value than 10.0 IU/L in our assay would have been used. This cutoff value already identified only 2.5% of the patients as having signs of a decreased ovarian reserve. A screening test identifying even fewer patients would inevitably have a severely reduced specificity and would become an irrelevant test because of the fact that the prevalence of positive test results would become too low. Also, as is shown in Figure 2 in our paper, the pregnancy rates in patients with basal FSH values over 15.0 and 20.0 IU/L were not significantly different from those in patients with basal FSH values over 10.0 IU/L.

In our opinion there is no direct need to correlate our assay to the Leeco Diagnostics assay for reasons of small study sample, because our study was conducted in a large cohort of subfertile pairs (n = 2,009). However, correlating the assays may further clarify our findings. Our assay was calibrated using the World Health Organization Second International Reference Preparation for FSH (78/549). Compared to the follow-up study by Scott et al. in patients from a general subfertility population (3), we used exactly the same cutoff value for basal FSH (10.0 IU/L). According to Scott et al., the cutoff value of 10.0 IU/L for assays calibrated against the Second International Reference Preparation for FSH compares to a cutoff value of 25.0 IU/L for the Leeco Diagnostics assay, calibrated against the Second International Reference Preparation for FSH (78/549)-hMG preparation (3).

In our article, the control group is being described as age-matched. We did not mention the fact that the size of the entire patient population (n = 2,009) also permitted us to match for the date of registration (matched within two months around the date of enrollment).

Finally, there is no reason for concern that our findings may have been influenced by the use of different protocols for improving ovarian response in "poor responders," because the use of high-dose gonadotropin treatment schemes has been shown to be ineffective in these patients (4, 5). Also, the majority of pregnancies in our elevated FSH group were spontaneous pregnancies.

Joris M. van Montfrans, M.D.
Annemieke Hoek, M.D., Ph.D.
Marcel H. A. van Hooff, M.D., Ph.D.
Cornelis B. Lambalk, M.D., Ph.D.
Research Institute for Endocrinology, Reproduction, and Metabolism
Department of Obstetrics and Gynecology
Vrije Universiteit Medical Center
Amsterdam, The Netherlands
November 19, 2000

References

1. van Montfrans JM, Hoek A, van Hooff MHA, de Koning CH, Tonch N, Lambalk CB. Predictive value of basal follicle-stimulating hormone