Value of the day 3 follicle-stimulating hormone measurement

Erin F. Wolff, M.D., and Hugh S. Taylor, M.D.

Department of Obstetrics, Gynecology, and Reproductive Sciences, Yale University School of Medicine, New Haven, Connecticut

Although elevated day 3 FSH is associated with diminished ovarian reserve, the predictive value is low in young women. Its use in this population as an exclusion criterion is unjustified. (Fertil Steril 2004;81:1486–8. ©2004 by American Society for Reproductive Medicine.)

In the setting of infertility, there are many ways to assess the probability of a successful pregnancy. It is important to accurately predict the probability of success of a treatment because of the substantial amount of time, money, and emotion that is invested by patients undergoing infertility treatment. Several tests of ovarian reserve have been introduced that may help predict IVF outcome; these include the clomiphene citrate challenge test, the GnRH-agonist stimulation test, measurements of basal inhibin B and antimullerian hormone (AMH), and quantitative ultrasound. However, most studies to date have looked at one test in particular, basal FSH, as a measure of ovarian reserve and IVF outcome (1). Day 3 FSH has been used routinely in the setting of IVF. Recently, the clinical utility of this test has been questioned.

The ovarian hormones estrogen and P are generally at their nadir in the late luteal or early follicular phase. This release of negative feedback on the hypothalamus results in an increase in GnRH pulse frequency. This in turn stimulates an increase in the circulating levels of FSH recruiting the next cohort of ovarian follicles (2). Reproductive senescence is typified by an accelerated decline in the follicular pool and by subfertility, as manifested by altered patterns of follicular growth and hormone secretion. The follicular phase may be shortened and the estrogen and inhibin levels altered.

Basal levels of FSH generally rise as the ovarian reserve decreases. For this reason, clinicians began using FSH in the evaluation of the infertile patient as a means of predicting ovarian reserve. The ability to make this prediction is important apart from age, because decreased ovarian reserve can occur over a range of ages. Elevated FSH levels are associated with lower pregnancy rates (3, 4). However, controversy exists as to whether this test is superior to age in assessing reproductive potential. Toner et al. (5) showed day 3 FSH to be an independent predictor of IVF outcome. Similarly, no differences in pregnancy or implantation were demonstrated between younger and older women with elevated day 3 FSH (6). In contrast, others have shown FSH to be of less value than age in predicting pregnancy rates (7), especially moderate elevations in younger women (8).

Day 3 FSH levels can be affected by many variables. A frequent problem that confounds day 3 FSH interpretation is sampling error. Appropriate timing of FSH testing presents a particular problem for one common group of infertile women—those with polycystic ovary syndrome (PCOS). These patients have longer, irregular cycles and are prone to anovulatory bleeding. Irregular cycles or spotting make it difficult to obtain an accurately timed sample. Even if using an appropriately timed collection, Lambalk and de Koning (9) reviewed evidence that intercycle variation, intersample variation (within assay and between assay), and hourly variations may result in disparate FSH measurements.

The existence of families with elevated FSH and increased frequency of dizygotic twinning
due to a hypothalamic-pituitary system with lower responsiveness to the ovarian hormone negative feedback suggests that genetic variation may account for elevated FSH levels in some young women (9). Another potential confounding factor suggested by Perez et al. (10) is genetic polymorphisms in the FSH receptor described in some young, fertile patients with elevated day 3 FSH levels. Normal reproductive variants like these, which result in elevated FSH measurements, may well be just as common as premature diminished ovarian reserve in younger patients.

Difficulty in interpreting day 3 FSH measurements occurs in older populations as well. The specificity of day 3 FSH in predicting pregnancy could be limited by the decrease in oocyte quality with age. This points to the importance of which outcome is used, namely, ovarian response or pregnancy rates. Ovarian response may be adequate, while the oocyte quality is not (1). A normal FSH does not ensure a successful pregnancy, especially in older patients.

When using any test in a given patient population, it is important to understand the potential value and limitations of the test. While the exact sensitivity and specificity of FSH measurements vary between studies and depend on the assay and reference values used, some generalizations can be made. The sensitivity, specificity, and likelihood ratio of the ability of day 3 FSH to predict nonpregnancy were recently compiled in a meta-analysis (1). As expected, the sensitivity rises when lower cutoff values are used at the expense of specificity. At higher cutoff levels, sensitivity falls as specificity increases. In fact, pregnancy rates appear to follow a continuum, with subtle declines noted from as low as 7 IU/L and rising thereafter.

As with any test, the sensitivity and positive predictive value of day 3 FSH rise as the incidence rises in the group studied. Infertility rates are higher in older patient populations than in younger ones. It follows that clinical performance of day 3 FSH improves in older patient populations. Given the conflicting data on the relevance of day 3 FSH, it is possible that we need to standardize how we use the test and in what populations.

Optimally, we should maximize the clinical usefulness of day 3 FSH measurements. It seems clear that day 3 FSH levels are not best suited for younger populations, largely because of the lower frequency of decreased ovarian reserve in this group. The statistical performance of this test is similar to another test used routinely in obstetrics, namely, the triple screen for Down’s syndrome, neural tube defects, and trisomy 18. The triple screen uses the profile of estradiol, hCG, and AFP in the maternal circulation to calculate the likelihood of abnormalities after adjustment for age. Triple screen results are reported as ratios and are based on the observation that these abnormalities occur in a higher frequency in an older population.

Can we use the triple screen test as a model for clinical application of day 3 FSH? The probability of poor ovarian response could be reported as likelihood ratios based on day 3 FSH levels, which are then adjusted for age. Advantages include that its format is familiar to both patients and clinicians. It gives patients concrete numbers on which to base very personal decisions. It is especially suited to tests that predict risk as a continuum and correlate with age. Perhaps test performance could be enhanced with the addition of inhibin levels or other hormones to create risk profiles for decreased ovarian reserve.

These data raise interesting questions regarding the clinical application of day 3 FSH levels. It is clear that neither day 3 FSH nor other currently available tests are 100% predictive of who will achieve pregnancy or even of ovarian reserve. Many argue that it is not a useful routine test before IVF. Given the test’s limitations, is it ethical to base inclusion/exclusion criteria for IVF on this test?

While many data exist on the usefulness of day 3 FSH, much of it is conflicting. It is clear that more accurate ways of predicting ovarian reserve are needed. Rigorous studies on the usefulness of newer tests like the clomiphene citrate challenge test, GnRH-agonist stimulation test, measurement of AMH (11, 12) and basal inhibin B (13, 14), and quantitative ultrasound will evaluate their clinical utility. However, at this time, it appears that exclusion of younger patients from IVF based on elevated day 3 FSH levels is not supported. It is important to recognize that this test is only a screening tool to predict ovarian reserve. Perhaps we can use day 3 FSH in a new and better way. This would include the use of alternate models for evaluating and reporting day 3 FSH, inclusion of additional measures of ovarian reserve, and adjustment for age. At the present time, there is little role for day 3 FSH evaluation in young healthy women.

References
8. Esposito MA, Coutifaris C, Barnhart KT. A moderately elevated day 3 FSH concentration has limited predictive value, especially in young women. Hum Reprod 2002;17:118–23.
10. Perez MM, Gromoll J, Brehm HM, Gassner C, Nieschlag E, Simoni M. Ovarian response to follicle-stimulating hormone (FSH) stimulation