Defining thresholds for abnormal premature progesterone levels during ovarian stimulation for assisted reproduction technologies

Micah J. Hill, D.O.,[a] Mae Wu Healy, D.O.,[a] Kevin S. Richter, Ph.D.,[b] Toral Parikh, M.D.,[a] Kate Devine, M.D.,[b] Alan H. DeCherney, M.D.,[a] Michael Levy, M.D.,[b] Eric Widra, M.D.,[b] and George Patounakis, M.D., Ph.D.[a]

[a] Program in Reproductive and Adult Endocrinology, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda; and [b] Shady Grove Fertility Reproductive Science Center, Rockville, Maryland

Objective: To evaluate methodologies to establish abnormal progesterone (P) levels on the day of trigger for recommending freeze only cycles.
Design: Threshold analysis and cost analysis.
Setting: Private ART practice.
Patient(s): Fresh autologous ART.
Interventions(s): None.
Main Outcome Measure(s): Live birth.
Result(s): Fourteen established statistical methodologies for generating clinical thresholds were evaluated. These methods were applied to 7,608 fresh ART transfer cycles to generate various P thresholds which ranged widely from 0.4 to 3.0 ng/mL. Lower thresholds ranged from 0.4 to 1 ng/mL and classified the majority of cycles as abnormal as well as required very large number needed to treat (NNT) to increase one live birth. Frozen embryo transfer was cost-effective when P was $\geq$ 1.5 ng/mL, with 12% of the population having an abnormal test result and an NNT of 13. Statistical and cost-effective thresholds clustered between 1.5 and 2.0 ng/mL.
Conclusion(s): Statistically significant thresholds for P were demonstrated as low as 0.4 ng/mL but resulted in a very large NNT to increase one live birth. A clinical benefit to a freeze-only approach was demonstrated above P thresholds ranging from 1.5 to 2.0 ng/dL. At these thresholds, elevated P has a demonstrable and clinically significant negative effect and captures a smaller percentage of the patient population at higher risk for fresh transfer failure, thus making freeze-only a cost-effective option. (Fertil Steril® 2018;110:671-9. ©2018 by American Society for Reproductive Medicine.)
El resumen está disponible en Español al final del artículo.

Key Words: Premature progesterone elevation, IVF, live birth, threshold analysis

There is now highly convincing observational data that premature progesterone (P) elevation on the day of trigger adversely affects assisted reproductive technology (ART) outcomes. Bosch et al. and Xu et al. separately published data from more than 14,000 ART cycles demonstrating that premature P elevations were negatively associated with the likelihood of pregnancy (1, 2). This negative association was confirmed in a meta-analysis of more than 60,000 ART cycles by Venetis et al. (3). Although not all studies confirm a negative association of premature P elevation and live birth, the majority of studies and the statistical synthesis of the available studies are convincing that this negative association occurs, resulting in decreased pregnancy and live birth in ART cycles (3).

Randomized controlled trials (RCTs) assessing a freeze-only strategy

Received March 21, 2018; revised April 19, 2018; accepted May 7, 2018.
M.J.H. has nothing to disclose. M.W.H. has nothing to disclose. K.S.R. has nothing to disclose. T.P. has nothing to disclose. K.D. has nothing to disclose. A.H.D. has nothing to disclose. M.L. has nothing to disclose. E.W. has nothing to disclose. G.P. has nothing to disclose.
Supported, in part, by the Intramural Research Program of the Program in Reproductive Endocrinology and Gynecology, National Institute of Child Health and Human Development, National Institutes of Health. The views expressed in this manuscript are those of the authors and do not reflect the official policy or position of the Department of the Army, Department of Defense, or the U.S. Government.
Reprint requests: Micah J. Hill, D.O., 10 Center Drive, Bethesda, MD 20876 (E-mail: hillmicah@mail.nih.gov).

Fertility and Sterility® Vol. 110, No. 4, September 2018 0015-0282/$36.00

https://doi.org/10.1016/j.fertnstert.2018.05.007

for premature elevated P are underway, and large retrospective data demonstrate that this approach ameliorates the negative association (4). The data are clear that premature P elevation is harmful, but there is no uniform definition to define premature P elevation. Fanchin et al. first published data in 1997 that demonstrated adverse effects on ART pregnancies when P was >0.9 ng/mL (5). Subsequently, most of the literature in the 2000s evaluated that threshold as a cutoff. Bosch et al. used a threshold of 1.5 ng/mL and Xu et al. used a threshold of 1.75 ng/mL (1, 2). We have published several papers using thresholds of both 1.5 and 2.0 ng/mL, clearly demonstrating a negative association of premature P elevation with live birth (4,6–10). The large meta-analysis from Venetis et al. clustered papers into threshold ranges and demonstrated that thresholds as low as 0.8–1.1 ng/mL resulted in decreased ART pregnancies (3). The 63 studies in that meta-analysis evaluated a wide range of thresholds from 0.4 to 3.0 ng/mL (3). In a recent *Fertility and Sterility* Journal Club Global, authors and experts on this topic debated threshold values, but were generally unwilling to define a specific numerical threshold at which they recommend a freeze-only approach (11).

The construction of thresholds to clinically define an abnormal test goes beyond simple statistical demonstration that the test results can differentiate patients based on outcomes. Changes in threshold values affect sensitivity, specificity, and positive and negative predictive values and have significant implications on the cost and number of patients that require further testing and intervention. In the case of premature P elevation, the test must accurately predict the absence of live birth while minimizing the costs associated with increased use of freeze-only. The threshold value selected should have strong positive predictive value, and the decision to freeze-only should be cost-effective at the chosen threshold.

The objective of the present study was to critically assess various methodologies for determining thresholds to define premature P elevation during ART stimulation. Threshold analyses and cost analyses were performed to determine optimal thresholds for defining premature elevated P and thereby create a clear counseling tool for patients for freeze-only of embryos versus fresh transfer.

MATERIALS AND METHODS

Study Design

This was a threshold analysis' study using retrospective cohort data of fresh ART cycles from 2013 to 2015. All fresh cycles were included if serum P was obtained on the day of trigger and a fresh embryo transfer occurred. This included all patients during the study time frame regardless of their ovarian response or stimulation protocol. The study was performed at Shady Grove Fertility Reproductive Science Center in Rockville, Maryland, with Institutional Review Board approval. Thresholds developed from this dataset were then validated with an earlier dataset in which P was measured on every patient on the day of trigger but no clinical decisions were made on the data (6).

Patients

All patients who underwent a fresh autologous embryo transfer with known serum P levels measured on the day of trigger were included in the analysis. Exclusion criteria included cycles where no embryo was transferred, donor-oocyte recipient cycles, frozen-thawed embryo transfers, and cycles without P measured on day of trigger. Otherwise, no exclusion criteria were applied.

Stimulation Protocol

Ovarian stimulation protocols included mixed FSH/hMG protocols with either GnRH agonist or GnRH antagonist for pituitary suppression (6). Oral contraceptive treatment was generally initiated 2–3 weeks before stimulation. For GnRH antagonist cycles, the antagonist was started when the lead follicle was 14 mm in size. For GnRH agonist cycles, 1 mg leuprolide acetate (Lupron) was administered daily subcutaneously during the last 3 days of oral contraceptive use. The leuprolide acetate was decreased to 0.25 mg when ovarian suppression was confirmed by means of ultrasound and serum E_2 <5 pg/mL. Ovarian stimulation was achieved with both FSH and hMG preparations. When two or more follicles were >18 mm in diameter, 10,000 IU hCG or 4 mg GnRH agonist was used for final oocyte maturation.

In general, patients predicted to be higher responders were placed on an antagonist protocol and were more likely to receive GnRH agonist trigger. Serum P levels were obtained on the day of trigger. Oocyte retrieval was performed 36 hours after the trigger injection. Fertilization was achieved with either conventional in vitro fertilization (IVF) or intracytoplasmic sperm injection as clinically indicated. After retrieval, patients triggered with hCG received vaginal P (100 mg Endometrin twice daily). GnRH agonist–triggered cycles also receiving supplementary low-dose hCG were prescribed vaginal P for luteal phase support in preparation for fresh embryo transfer. If the supplementary low-dose hCG was withheld (because of very high oocyte yield), GnRH agonist–triggered cycles received 50 mg intramuscular P in oil daily for more aggressive luteal support.

Ultrasound-guided embryo transfer was performed on day 3 or 5 if an adequate number of high-quality embryos were available. Embryos were graded as good, fair, or poor according to the simplified Society for Assisted Reproductive Technology scoring system (12). Serum hCG levels were assessed at 4 weeks' gestational age followed by ultrasound confirmation of an intrauterine pregnancy in all pregnant patients.

Serum P levels were measured with the use of a solid-phase competitive chemiluminescent enzyme immunoassay (Immunolyte 2000 progesterone assay; Siemens Medical Solutions Diagnostic). The lower limit of detection for the assay was 0.2 ng/mL, and the analytical sensitivity of the assay was 0.1 ng/mL. Intra-assay and interassay coefficients of variances were 6.7% and 7.2%, respectively. Previous data have demonstrated the utility of this assay in predicting ART live birth (4, 6–10).