FERTILITY AND STERILITY
Vol. 62, No. 1, July 1994

Printed on acid-free paper in U. S. A.

Luteal phase defect: the sensitivity and specificity of diagnostic methods in common clinical use*

John Jordan, M.D.
Kristin Craig, B.S.
Donald K. Clifton, Ph.D.
Michael R. Soules, M.D.†

Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, University of Washington, Seattle, Washington

Objective: To assess the sensitivity and specificity of common clinical tests used for the diagnosis of luteal phase defect (LPD).

Design: The sensitivity and specificity of these tests for predicting low integrated P levels over the luteal phase were calculated.

Setting: Outpatient reproductive endocrinology and infertility clinic at a university medical center.

Patients: Fifty-eight strictly defined normal women were used to determine normal integrated luteal phase P levels. The study population was a separate 34 women who either were normal (n = 15) or were being evaluated for infertility or recurrent abortion (n = 19). These 34 study subjects all had the following tests performed in the same menstrual cycle: daily reproductive hormone levels, daily assessment of preovulatory follicle size, late luteal endometrial biopsies, and BBT charts.

Main Outcome Measures: Basal body temperature, maximum preovulatory follicle size, dated endometrial biopsies, and serum P levels (single and multiple) were used in an attempt to predict which patients had low integrated P levels.

Results: Unacceptably low sensitivity and/or specificity levels were found for the following tests: appearance of BBT charts, luteal phase length, and preovulatory follicle diameter. Timed endometrial biopsy was found to have marginally acceptable sensitivity and specificity levels whether dated by next menstrual period or midcycle events. The best test for the prediction of low integrated P was a single serum P level from the midluteal phase that was <10 ng/mL (31.8 nmol/L) or a sum of three random serum P measurements that was <30 ng/mL (95.4 nmol/L) (also obtained in the midluteal phase).

Conclusions: Luteal phase defect is a relatively uncommon but important cause of infertility and/or habitual abortion. The recommended test for the determination of LPD is a midluteal phase single serum P level <10 ng/mL or the sum of three serum P levels that is <30 ng/mL. The endometrial biopsy is a second line test that is only recommended when LPD needs to be evaluated in a treated cycle (ovulation induction or supplemental P). Fertil Steril 1994;62:54–62

Key Words: Luteal phase defect, diagnostic tests, analysis, statistics

The clinical entity of luteal phase defect (LPD) is present when there is recurrent postovulatory deficiency in the production and/or effect of P from the corpus luteum (CL), leading to infertility and/or habitual abortion (1). There are a number of hor-

Received October 14, 1993; revised and accepted February 23, 1994.

* Supported by grants RO1 HD 18967 (M.R.S.) and MO1 RR00037 (Clinical Research Center, University of Washington, Seattle, Washington).

† Reprint requests: Michael R. Soules, M.D., Division of Reproductive Endocrinology and Infertility, 4225 Roosevelt Way NE, Suite 305, Seattle, Washington 98105 (FAX: 206-685-7818).

mones secreted by the CL in the luteal phase, including P, E2, 17-hydroxyprogesterone, relaxin, and inhibin. In terms of physiological and clinical relevance, the focus has been on P secretion and effect. In the majority of women with LPD, the primary problem appears to be an inadequate quantity or duration of P secretion (2). A small minority of cases of LPD are caused by reduced effectiveness of P, primarily on the endometrium. This report will focus on LPD as a disorder of P secretion. It should be noted that LPD does not stem from a single etiology; multiple causes funnel through this common pathway of decreased P secretion or effect (1).

Although LPD has been clearly described in research settings, the diagnosis, prevalence, and treatment of this syndrome in the clinical setting remains controversial. The currently available diagnostic tests for LPD are cumbersome and have not been properly validated. The original description of LPD in 1949 incorporated BBT charts, urinary pregnanediol levels, and endometrial biopsy as diagnostic tests (3). Additional diagnostic tests for LPD have been proposed and clinically incorporated since 1949, including single and multiple serum P levels, luteal phase length measured by various methods, preovulatory follicle size as determined by ultrasound (US), and timed endometrial biopsy (4). For true accuracy and clinical validity, each of these diagnostic methods needs to have been validated in a clinical population who were determined to have LPD by a recognized gold standard.

However, in the clinical literature, various investigators (4) have attempted to evaluate LPD diagnoses by comparing one diagnostic test with another. By consensus, the most common standard used is timed endometrial biopsy. However, the use of timed endometrial biopsy is an ideal example of the cyclic validation that has occurred throughout the history of LPD studies. When Noyes et al. (5) first described the criteria for interpreting timed endometrial biopsies, their results were based on BBT charts. Since this original validation in 1950, timed endometrial biopsy has become more accepted and is now being used to determine the efficacy of BBT charts in relation to LPD (6). To date, there are no completely validated diagnostic tests for LPD that are practical in the clinical setting (note that in the research setting, an integrated luteal P level is considered to be a relative gold standard, but obtaining daily serum P levels in the clinical setting is not practical). Much of the controversy that surrounds LPD can probably be traced to a lack of accuracy and sensitivity inherent in established (but not validated) diagnostic tests for LPD.

In this investigation, we have compiled and analyzed data collected during the course of several studies in which all the clinical tests that have been purported to diagnose LPD have been performed within the same menstrual cycle. These tests include the following: BBT chart, preovulatory US, timed endometrial biopsy, and daily serum P levels. We compared these tests with a relatively precise and reproducible standard of CL function-integrated P levels in serum over the luteal phase. In this study, it was our intent to determine which of these common clinical tests are the most sensitive and specific in predicting low integrated P levels.

MATERIALS AND METHODS

Individual menstrual cycle data from 34 women were examined from several studies performed recently at the University of Washington Medical Center, Seattle, Washington. All of these studies were approved by the Human Subjects Review Committee. In these studies, all of the tests in question were performed during the same cycle in two groups of women: a normal group of volunteers (n = 15) or women who were being evaluated for infertility or recurrent abortion (n = 19) (Table 1). To be included in the normal group, these women met the following criteria: [1] 18 to 35 years of age; [2] history of regular menstrual cycles; [3] weight within ±10% of ideal body weight (Metropolitan Life Tables, 1980); [4] biphasic BBT chart with a luteal phase length >12 days; [5] aerobic exercise <10 hours/wk; [6] known to have a regular dietary history; [7] normal reproductive history; and [8] not taking any medications on a regular basis. We performed four tests on all of the women.

Basal Body Temperature

Basal body temperature was measured daily by the women and recorded on standard temperature charts. Ovulation was considered to have occurred when there was a temperature rise of ≥0.4°C for 7 or more days immediately preceding a menstrual period. These charts were then blindly reviewed by three reproductive endocrinologists who were asked to estimate the length of the luteal phase, determine whether LPD was present by subjective criteria, and provide reasons for their diagnoses. All three physicians classified the patients with luteal

Table 1 Study Population

Subject	Age	Diagnosis*	Follicular phase length	Luteal phase length	Maximum preovulatory follicle size†	Integrated P
			d		*mm*	*(ng·day/mL)*
A	32	N	16	14	22.3	159
B	29	N	23	13	25	108
C	23	N	14	12	19.6	135
D	21	Infertility	14	12	19	97
E	32	Infertility	10	10	21.3	65
F	34	Infertility	12	12	23.6	100
G	36	Infertility	12	12	20	73
H	31	N	13	13	22.6	60
I	33	N	14	14	20.6	93
J	32	Habitual abortion	10	10	24.6	143
K	27	Infertility	11	11	24	111
L	35	Infertility	13	13	20.6	84
M	28	N	12	12	24.6	162
N	24	N	14	14	19.6	147
O	25	N	14	15	22	148
P	32	N	11	16	21	124
Q	25	N	14	12	25	92
R	24	N	15	11	22	164
S	37	Infertility	17	14	16.5	74
T	28	N	20	16	17.5	158
U	41	Habitual abortion	12	13	17.5	77
V	38	Infertility	15	13	16	142
W	33	N	14	10	24.5	102
X	39	Infertility	16	12	23.3	170
Y	34	Infertility	14	14	21	138
Z	39	Infertility	12	14	21	199
AA	31	Infertility	17	12	15	155
BB	30	Infertility	17	14	18.5	130
CC	35	Infertility	15	12	23.5	123
DD	37	Infertility	9	15	19	120
EE	25	N	18	11	22.5	168
FF	37	Infertility	12	14	24	64
GG	28	N	16	12	21.5	73
HH	32	Habitual abortion	27	13	22.6	120

* N, normal.

† Values are means.

lengths ≤11 days as having LPD. For the purpose of this study, a particular BBT chart was considered to be indicative of LPD when two of the three physicians made this diagnosis.

Follicle Size

An Acuson (Mountain View, CA) model 128 sector US scanner with a 5-mHz abdominal transducer was used to obtain daily estimates of mean follicle diameter throughout the follicular phase. The day of ovulation was indicated by the presence of at least two of the following findings: an acute decrease in follicle diameter, an abrupt increase in free intraperitoneal fluid, and/or the appearance of intrafollicular echoes. For purposes of this study, a maximum mean preovulatory follicle size of <17 mm was considered to indicate LPD (7, 8).

Endometrial Biopsies

Biopsies were performed in the late luteal phase of each study cycle with either a Novak curette or a Pipelle (Unimar, Inc., Wilton, CT) sampling device. These biopsies were assigned menstrual cycle dates according to the criteria of Noyes et al. (5) by two pathologists, each of whom read all the slides in a blind fashion at the completion of the study. The slides of the patients (subjects) were in a random order in each set reviewed by each pathologist. There was initial agreement between the pathologists within ±1 day for >90% of the slides. The two pathologists reviewed together those slides in which there was a discrepancy of >1 day and reached a consensus while still blinded to the other study parameters. Biopsies were considered out of phase if they were dysynchronous by >2 days in

relation to either the 1st day of the next menstrual period, the day of the LH surge as determined by RIA, or ovulation as determined by US examination performed at midcycle.

Serum P Measurements

Venous blood samples were collected between 8:00 A.M. and 10:00 A.M. daily from the onset of menstruation to the 1st day of the next menstrual period; serum was separated, frozen, and stored at −4°C. Serum P was measured by RIA using reagents supplied by Diagnostic Products Corporation (Los Angeles, CA). The serum P results are expressed in ng/mL (conversion factor to SI units, 3.18). Serum LH was measured by double-antibody RIAs using reagents supplied by the National Institute of Health with LER-907 as the reference preparation. The methodology of both assays has been described previously (2). The intra-assay and inter-assay coefficients of variation for the P and LH assays were 9.6% and 13.2% for P and 5.5% and 8.4% for LH. The peak daily LH level at midcycle was considered to be the day of the LH surge. The luteal phase was defined as the day after the LH surge (day +1) up to and including the day before the next menstrual period.

Single and multiple midluteal phase P levels have frequently been used to diagnose LPD (9–12). To evaluate single and multiple P determinations as diagnostic criteria, we used a computer-assisted procedure to randomly select samples from the daily luteal phase P values that we obtained from each subject. A relatively narrow span (+5 to +9) and a somewhat wider span (+4 to +11) of days relative to the LH surge were selected for sampling as two clinically applicable demarcations of the midluteal phase (10). Each subject had a single serum P value randomly selected between days 5 and 9 and another between days 4 and 11 of the luteal phase in her study cycle. Furthermore, each subject had three serum P levels randomly selected between days +5 to +9, and the sum of these three P levels was determined. We also repeated the random selection of P measurements (single and multiple) four times in an attempt to ascertain sampling variance.

Luteal integrated P levels were determined by summing daily serum P levels, starting with the day after the LH surge and ending with the day before the next menstrual period. (Integrated P is an estimate of total P output over the luteal phase.) Because we consider integrated P to be the best estimate of CL function available, we used this measurement as the standard by which we evaluated the other tests (BBT, random P levels, US, endometrial biopsy). To determine the normal range for integrated P, we used integrated P values measured in a separate pool of normal women (n = 58) that met the criteria listed above for normal volunteers. The prevalence of LPD in the general population has been estimated to be 10% (4); therefore, because 10% of the women in this pool had integrated P levels <80 ng·days/mL (254.4 nmol·days/L), we defined LPD as an integrated P value <80 ng·days/mL. Previously published studies (13–15) have used other criteria to define low integrated P levels. Based on data from 13 normal women, del Pozo et al. (13) derived a value of 107 ng·days/mL (340.26 nmol·days/L) as the lower limit of normal integrated P. Similarly, Wu and Minassian found in a group of 13 normal women with endometrial biopsies >2 days out of phase that the integrated P was 113 ± 8.5 ng·days/mL (27.03 nmol·days/L) as compared with 170 ± 8.3 ng·days/mL (26.39 nmol·days/L) in a group of 39 normal infertile women (14). Schweiger and colleagues (15) in a study of 18 female athletes in whom a high prevalence of LPD was diagnosed found an integrated P level of 100 ng·days/mL (318 nmol·days/L) to be the critical cutoff. Our figure of 80 ng·days/mL for integrated P, which was derived from a much larger data base, is more conservative than the values determined in these previous studies.

We used this criterion (integrated P <80 ng·day/mL) to compare the results of the various tests and combination of tests to assess their sensitivity and specificity in predicting low integrated P levels. In this context, sensitivity is the probability that the test will detect low luteal phase integrated P levels when it is present; specificity is the probability that the test will be negative when integrated luteal phase integrated P levels are normal.

RESULTS

The 34 women in this study ranged in age from 21 to 41 years. Seven (21%) of the women had LPD, as determined by an integrated P of <80 ng·days/mL. A summary of our evaluation of each test compared with integrated P can be found in Table 2. The appearance of the BBT chart, even to expert observers, was a very insensitive predictor of LPD. Likewise, luteal phase length, whether it was estimated by BBT chart or calculated after determin-

Table 2 Performance of Various Tests in Predicting Deficient Luteal P Secretion*

Test	Sensitivity	Specificity
Appearance of BBT chart†	14	74
Luteal phase length <11 days as determined by:		
BBT	0	82
LH surge	14	82
US	14	74
Preovulatory follicle diameter <17 mm	14	93
Out-of-phase timed endometrial biopsy‡ dated according to:		
Next menstrual period	57	44
US	29	56
LH surge	29	56
Single random serum P measurement§		
<5 ng/mL‖		
5 to 9 days after ovulation	14	98
4 to 11 days after ovulation	11	98
<10 ng/mL		
5 to 9 days after ovulation	86	83
4 to 11 days after ovulation	82	81
<15 ng/mL		
5 to 9 days after ovulation	100	43
4 to 11 days after ovulation	100	41
Sum of three random serum P measurements§		
<30 ng/mL	100	80
<25 ng/mL	78	92
<20 ng/mL	50	99
<15 ng/mL	0	100
Combined out-of-phase timed endometrial biopsy per single P <10 ng/mL		
Both positive	71	93
Either positive	100	63

* Based on 80 ng·day/mL.
† Two of three physicians made diagnosis of LPD based on rate, magnitude, and/or pattern of temperature rise or drop.
‡ >2 days out of phase.
§ Average of four random trials (Table 3).
‖ Conversion factor to SI units, 3.18.

ing the exact day of the LH surge or follicle disappearance, was very insensitive as well. Another test with very low sensitivity was preovulatory follicle size, with only 14% of those women with low integrated P being correctly identified. No subject had the luteinizing unruptured follicle syndrome, that is, persistence and/or growth of a follicle beyond the day of ovulation as identified by the onset of P secretion.

Timed endometrial biopsy was somewhat more sensitive, yet both sensitivity and specificity were only approximately 50% at best (Table 2). Timed endometrial biopsy appeared to be less sensitive when the biopsies were dated according to midcycle events (LH surge and ovulation) compared with dating by next menstrual period. There were 11, 8, and 6 out of phase timed endometrial biopsies according to the next menstrual period, LH surge, and US criteria, respectively. The next menstrual period method was more sensitive (correctly identified 5 of 7 women with low integrated P) but less specific (falsely identified 6 women with normal integrated P) than dating by midcycle events.

The results of our four random trials selecting P level on particular cycle days are found in Table 3. The specificity and sensitivity results presented in Table 2 are averages obtained from the four trials in Table 3. The best results for single P measurements were those taken either between days 5 and 9 or days 4 and 11 after ovulation and using <10 ng/mL as the criteria. In the majority of the trials, these measurements had sensitivities approaching 100%; therefore, most women with integrated P levels <80 ng·day/mL were correctly identified as having LPD. When the <5 ng/mL (15.9 nmol/L) criterion was used, most of the women with low integrated P were missed. Conversely, use of the <15 ng/mL (47.7 nmol/L) criterion resulted in a high rate of false LPD diagnoses (a specificity of approximately 40%). The use of three random samples obtained between cycle days +5 to +9 had similar results compared with the single sample results but demonstrated more consistency from trial to trial. The best criteria for the sum of three measurements was <30 ng/mL with a sensitivity of 100% and a specificity of 70% to 93%. Although the use of a sum of

Table 3 Four Trials of Single and Multiple Random P Samples

Trial	Sensitivity/specificity 1	2	3	4
	%			
Single samples				
<5 ng/mL*				
Days 5 to 9	14/100	29/93	14/100	0/100
Days 4 to 11	14/100	14/93	14/100	0/100
<10 ng/mL				
Days 5 to 9	100/85	57/85	86/81	100/82
Days 4 to 11	100/82	86/74	71/93	71/74
<15 ng/mL				
Days 5 to 9	100/37	100/33	100/44	100/56
Days 4 to 11	100/48	100/33	100/41	100/41
Sum of three samples†				
<30 ng/mL	100/78	100/70	100/93	100/78
<25 ng/mL	71/93	100/81	71/93	71/100
<20 ng/mL	57/96	42/100	43/100	57/100
<15 ng/mL	0/100	0/100	0/100	0/100

* Conversion factor to SI units, 3.18.
† Obtained between days 5 and 9 after LH surge.

<15 ng/mL has been proposed in the LPD literature (9), decreasing the criteria <30 ng/mL appears too low to correctly identify low integrated P, as illustrated by the sensitivity levels for <25, <20, and <15 ng/mL, respectively (<79.5, <63.6, and <47.7 nmol/L, respectively).

Last, we also tried the combination of timed endometrial biopsy based on next menstrual period (>2 days out of phase) and a single P measurement (<10 ng/mL) taken between days 5 and 9. By combining these tests and considering either test positive equivalent to a diagnosis of LPD, we increased our sensitivity but also decreased the specificity. If we required that both tests be positive, we greatly decreased the sensitivity, but it did not increase the specificity.

DISCUSSION

Although a number of clinical tests have been proposed and are currently used to diagnose LPD, they all suffer from a lack of adequate validation and standardization. Some of the deficiencies associated with the validation of these tests are as follows: [1] Almost all of the tests have been evaluated in relation to prior tests but not against accepted physiological criteria. [2] Some reports have used the successful treatment of an infertile patient who appears to have LPD as diagnosed by a given test as an indication that the woman initially had LPD as diagnosed by that particular test (16–19). Obviously, the pregnancy in question could have occurred by chance or by simultaneous treatment of other conditions. [3] Some have compared results for different tests during different cycles of the same woman, ignoring the well-recognized fact that CL function varies from cycle to cycle. [4] Sometimes the results have been compared with a particular test because it has been used commonly in the LPD literature and/or it appears to be a good test because it is a bioassay (e.g., timed endometrial biopsy) (6, 10). To provide an adequate evaluation of the currently used tests, a reproducible and physiologically relevant standard is necessary for comparison, and the comparison must be done with results that are obtained during the same cycle.

Because all of the existing LPD tests are based on the secretion of P, we chose to use integrated P as our standard for comparison. The measurement of serum P is relatively precise and reproducible. Because P is secreted in a pulsatile fashion (20, 21), P values can fluctuate dramatically from sample to sample; however, by integrating P values over the entire luteal phase, the effect of this variation is reduced dramatically. It should be kept in mind that the use of integrated P does assume that LPD is primarily a result of P secretion rather than a problem with the end organ effect of P. Integrated luteal P as calculated from daily hormone measurements is the most accepted standard for determining LPD in the research literature (22–26), but the cost and inconvenience of obtaining daily serum samples makes this approach clinically impractical.

Timed endometrial biopsy appears to be a good initial choice as a diagnostic test because it measures the effect of P on the end organ, thus eliminating the need to determine whether LPD represents a deficiency in P production or P effect. Yet timed endometrial biopsy as a diagnostic test has a number of inherent flaws as follows: [1] it is dependent on a subjective histologic interpretation; [2] different investigators have used different criteria for interpretation (e.g., the handling of glandular-stromal disparity); and [3] only moderate reproducibility of readings exist, even when the same specimen is read several times by a single pathologist (27). Timed endometrial biopsy has been validated as the definitive test for LPD by comparing its results to unproven criteria such as BBT chart and single P measurements (5, 28). There has been some disagreement over whether to use a 2-day, or >2-day, out-of-phase criterion to indicate LPD (29). There have also been reports in the literature that timed endometrial biopsy is more accurate if the histologic dates are compared with what are assumed to be more precise midcycle events (27, 30). We found the performance of timed endometrial biopsy as a predictor of low integrated P to be relatively poor, regardless of whether 2 or >2 days out-of-phase criterion was used, or whether it was based on midcycle events or next menstrual period. In fact, next menstrual period appeared to be superior (more sensitive) to dating by midcycle events in our study. Although timed endometrial biopsy has been well accepted in the literature, our data suggest that timed endometrial biopsy does not correlate well with low integrated P, and we would not recommend it as the standard test for the clinical diagnosis of LPD. However, it still remains the test of choice to evaluate an ovulation induction cycle for LPD because adequate integrated P and corresponding random serum P levels have not been determined for stimulated cycles. Likewise, when LPD is treated with supplemental P, an endometrial biopsy would be recommended to determine adequacy of therapy because serum P levels would

vary tremendously with the timing of exogenous P administration.

Other popular tests (BBT chart, luteal length, follicle diameter) in current clinical usage for the diagnosis of LPD were also evaluated. The sensitivity and/or specificity of these other tests were all even lower than timed endometrial biopsy. Basal body temperature charts were the first standard used for the diagnosis of LPD (3). Presently, BBT charts are used both to determine luteal length and to diagnose LPD by subjective criteria. Downs and Gibson (6) in a study of 40 women, 20 with in-phase and 20 with out-of-phase biopsies, showed no difference in the rate of temperature rise between the two groups, although there was a shorter luteal length in a few of the LPD subjects. Our results concur with this prior study with sensitivity and specificity levels that indicate the BBT chart is an inferior method for the diagnosis of LPD. Luteal length is a relatively easy measure of luteal function. Lenton and colleagues (31) found luteal lengths of 10 days or less were abnormal when a relatively large population was studied. In another study, Smith et al. (32), in their study of 95 women with unexplained infertility and 29 controls, found no difference in occurrence of luteal phase $\leq$ 11 days (9% and 8%, respectively). Our data showed that luteal lengths as calculated by three different methods was an insensitive test for LPD and perhaps reflects only the most severe cases. Certainly, women with low integrated P levels have been found to have significantly shorter luteal phase lengths. However, luteal phase length does not appear to be a good diagnostic test for LPD.

Ultrasound assessment of follicle diameter is a popular test in the clinical setting. Geisthoval et al. (8) reported a smaller mean follicle diameter in LPD cycles compared to normal cycles (17.7 ± 2.9 vs 23 ± 2.3 mm). Likewise, Check and colleagues (7) have found US to be useful in the diagnosis of LPD, noting that 40% of women with two out-of-phase biopsies had consistent small follicle size (<17 mm) and an additional 12% with an occasional small follicle. However, we reported previously in 10 women with low integrated P levels and out-of-phase biopsies in a study cycle that there were no decreases in mean follicle diameter (2). In the present study, a maximum preovulatory follicle size of 17 mm or less was unacceptably insensitive for the diagnosis of low integrated P levels. It is concluded that decreased follicle diameter is not a common finding in LPD cycles, and US assessment of follicle size is not a recommended test for the diagnosis of LPD.

The determination of P levels in single or multiple blood samples frequently has been recommended as a diagnostic test for LPD. Because of the known pulsatile nature of P secretion (5 to 12 pulses/d with no standard diurnal pattern), the analysis of P in single or multiple samples is of questionable efficacy (20, 21). This is probably the reason that the range of suggested criteria for adequate P levels is quite large. For single P determinations, levels as low as 5 ng/mL and as high as 15 ng/mL have been suggested as proper criteria for single samples (11, 33). In regards to the summation of three samples, levels of 10 and 15 ng/mL have been suggested (9). Taking into account the secretory pattern of P and the large variation in cutoff levels in the literature, we analyzed several levels as indicative of low integrated P in this study. The present study demonstrated that serum P levels combined the highest sensitivity and specificity and thus were more accurate at determining low integrated P levels. The best results occurred with the use of a single blood sample with a P level of <10 ng/mL or a sum of three samples with a P level of <30 ng/mL, with all samples obtained between days 5 and 9 after ovulation. In examining other cutoff criteria for single samples, the use of <5 ng/mL appeared to be too stringent, missing most of the women with LPD, and a cutoff criteria of <15 ng/mL appeared to be too liberal and resulted in many false-positives. Although the predictive value for low integrated P levels obtained with a single sample were frequently equal with the three sample criteria, the use of three samples was more consistent in achieving high sensitivities (Table 3).

Based on the results of this study, we currently are recommending that three serum P levels be obtained between cycle days 5 and 9 after ovulation. If these levels add up to <30 ng/mL in two spontaneous menstrual cycles, then it would appear that persistent low integrated P levels are present, and a diagnosis of LPD would be reasonable. The use of luteal phase blood sampling for P would require the identification of the day of ovulation to time luteal phase sampling. Several methods could be used including BBT chart, US, daily LH measurements, and urinary LH kits. Daily blood samples for serum LH levels and US are expensive and inconvenient for patients. Basal body temperature charts are too inaccurate for pinpointing the day of the LH surge. Urinary LH tests can be done easily at home by the patient and have been found to be quite accurate. If three blood samples are obtained between luteal phase days 5 to 9, for cost-effectiveness, equal ali-

quots of serum from each sample could be pooled with a single mean P level of 10 ng/mL used as the criteria. For the tests we evaluated in this study, the current charges at our institution (including the visit to obtain the samples) are as follows: [1] daily US examinations (3), $198.00; [2] endometrial biopsy, $272.00; and [3] serum P, three samples pooled, plus urine LH kit, $87.00.

We used <80 ng · day/mL as the criteria for low integrated P in this study. Out of interest, we recalculated our sensitivity and specificity based on 100 ng · day/mL as our standard because that level had been reported as a common criteria in prior studies (2). The results of the various clinical tests, including single and multiple blood levels for serum P, showed no apparent differences in sensitivity and specificity when 100 ng · day/mL was used as the standard.

A relevant question pertains to the applicability of these findings to a general (wider) infertility population. This study was performed in a selected and relatively small population with the recognition that it is difficult to obtain similar data in a large group of women. It has been estimated that 5% to 10% of the referral infertility population has LPD. The population studied here had a 21% incidence of LPD when LPD is equated with low integrated P. The variance in randomly selecting endocrine data was estimated in this population (Table 3). It is unlikely that furthering sampling trials in this population would significantly improve our estimates of the sensitivity and specificity for these tests as applied to the general population. Although we would expect a prevalence of LPD (low integrated P) in a general population to be <21%, we have no reason to believe that the sensitivity and specificity of the tests we evaluated would be affected by a difference in prevalence.

It is relatively easy to increase luteal phase P levels in women with apparent LPD by treatment with P supplementation or clomiphene citrate. Such treatment has very few side effects. Therefore, it seems clinically more valuable to risk treating some women without LPD (false-positive) to treat all those who may have LPD. Thus, the test with the highest sensitivity, those that miss the fewest number of truly positive women, would appear to be the best test. From the study, it is apparent that the sum of three serum P levels <30 ng/mL obtained between cycle days +5 to +9 is the best current diagnostic test for LPD for use in the clinical setting.

Acknowledgments. The authors thank Leon R. Spadoni, M.D., and Donald E. Moore, M.D., (Department of Obstetrics and Gynecology, University of Washington, Seattle, WA) for reviewing BBT charts; Mr. Arlen Sarkissians, Ms. Dorothy McGuinness, and Ms. Florida Flor (Department of Veterans Affairs Medical Center, Seattle, WA) for laboratory expertise; Marit Ek, M.D., (Department of Pathology, University of Washington, Seattle, WA) for reading endometrial biopsy slides; Ms. Gretchen S. Davis for helping to coordinate this research project; and Mr. Jesse T. Morelli (University of Washington, Seattle, WA) for secretarial assistance.

REFERENCES

1. March CM, Shoupe D. Luteal-phase defects. In: Mishell DR, Davajan V, Lobo RA, editors. Infertility, contraception and reproductive endocrinology. Boston: Blackwell Scientific Publications, 1991:793–806.
2. Soules MR, McLachlan RI, Ek M, Dahl KD, Cohen NL, Bremner WJ. Luteal phase deficiency: characterization of reproductive hormones over the menstrual cycle. J Clin Endocrinol Metab 1989;69:804–12.
3. Jones GES. Some newer aspects of the management of infertility. JAMA 1949;141:1123–9.
4. McNeely MJ, Soules MR. The diagnosis of luteal phase deficiency: a critical review. Fertil Steril 1988;50:1–15.
5. Noyes RW, Hertig AT, Rock J. Dating the endometrial biopsy. Fertil Steril 1950;1:3–25.
6. Downs KA, Gibson M. Basal body temperature graph and the luteal phase defect. Fertil Steril 1983;40:466–80.
7. Check JH, Goldberg BB, Kurtz A, Adelson HG, Rankin A. Pelvic sonography to help determine the appropriate therapy for luteal phase defects. Int J Fertil 1984;29:156–8.
8. Geisthoval F, Skubsch U, Zabel G, Schillinger H, Breckwoldt M. Ultrasonographic and hormonal studies in physiologic and insufficient menstrual cycles. Fertil Steril 1983;39:277–83.
9. Abraham GE, Maroulis GB, Marshall JR. Evaluation of ovulation and corpus luteum function using measurements of plasma progesterone. Obstet Gynecol 1974;44:522–5.
10. Hensleigh PA, Fainstat T. Corpus luteum dysfunction: serum progesterone levels in diagnosis and assessment of therapy for recurrent and threatened abortion. Fertil Steril 1979;32:396–400.
11. Radwanska E, Swyer GIM. Plasma progesterone estimation in infertile women and in women under treatment with clomiphene and chorionic gonadotropin. J Obstet Gynaecol Br Commonw 1974;81:107–12.
12. Johansson EDB. Progesterone levels in peripheral plasma during the luteal phase of the normal human menstrual cycle measured by a rapid competitive protein binding technique. Acta Endocrinol (Copenh) 1969;61:592–606.
13. del Pozo E, Wyss H, Tolis G, Alcaniz J, Campana A, Naftolin F. Prolactin and deficient luteal function. Obstet Gynecol 1979;53:282–5.
14. Wu CH, Minassian SS. The integrated luteal progesterone: an assessment of luteal function. Fertil Steril 1987;48:937–40.
15. Schweiger U, Laessle RG, Tuschl RJ, Broocks A, Krusche T, Pirke KM. Decreased follicular phase gonadotropin secretion is associated with impaired estradiol and progesterone secretion during the follicular and luteal phases in nor-

mally menstruating women. J Clin Endocrinol Metab 1989;68:888–92.
16. Balasch J, Vanrell JA, Marquez M, Burzaco I, Gonzalez-Merlo J. Dehydrogesterone versus vaginal progesterone in the treatment of the endometrial luteal phase deficiency. Fertil Steril 1982;37:751–4.
17. Daly DC, Walters CA, Soto-Albors CE, Riddick DH. Endometrial biopsy during treatment of luteal phase defects is predictive of therapeutic outcome. Fertil Steril 1983;40:305–10.
18. Murray DL, Reich L, Adashi EY. Oral clomiphene citrate and vaginal progesterone suppositories in the treatment of luteal phase dysfunction: a comparative study. Fertil Steril 1989;51:35–41.
19. Huang KE. The primary treatment of luteal phase inadequacy: progesterone versus clomiphene citrate. Am J Obstet Gynecol 1986;155:824–8.
20. Filicori M, Butler JP, Crowley WF. Neuroendocrine regulation of the corpus luteum in the human: evidence for pulsatile progesterone secretion. J Clin Invest 1984;73:1638–47.
21. Soules MR, Clifton DK, Steiner RA, Cohen NL, Bremner WJ. The corpus luteum: determinants of progesterone secretion in the normal menstrual cycle. Obstet Gynecol 1988;71:659–66.
22. Loucks AB, Mortola JF, Girton L, Yen SSC. Alterations in the hypothalamic-pituitary-ovarian and the hypothalamic-pituitary-adrenal axes in athletic women. J Clin Endocrinol Metab 1989;68:402–11.
23. Pirke KM, Schweiger U, Strowitzki T, Tuschl RJ, Laessle RG, Broocks A, et al. Dieting causes menstrual irregularities in normal weight young women through impairment of episodic luteinizing hormone secretion. Fertil Steril 1988; 51:263–8.
24. Stouffer RL, Hodgen GD. Induction of luteal phase defects in rhesus monkeys by follicular fluid administration at the onset of the menstrual cycle. J Clin Endocrinol Metab 1980;51:669–71.
25. Hutchison J, Zelenzik AJ. The corpus luteum of the primate mentrual cycle is capable of recovering from a transient withdrawal of pituitary gonadotropin support. Endocrinology 1985;117:1043–9.
26. McLachlan RI, Cohen NL, Vale WW, Rivier JE, Burger HG, Bremmer WJ, et al. The importance of luteinizing hormone in the control of inhibin and progesterone secretion by the human corpus luteum. J Clin Endocrinol Metab 1989;68:1078–85.
27. Li TC, Rogers AW, Lenton EA, Dockery P, Cooke I. A comparison between two methods of chronological dating of human endometrial biopsies during the luteal phase, and their correlation with histologic dating. Fertil Steril 1987;48:928–32.
28. Annos T, Thompson IE, Taymor ML. Luteal phase deficiency and infertility: difficulties encountered in diagnosis and treatment. Obstet Gynecol 1980;55:705–10.
29. Davis OK, Berkeley AS, Naus GJ, Cholst IN, Freedman KS. The incidence of luteal phase defects in normal, fertile women, determined by serial endometrial biopsies. Fertil Steril 1989;51:582–6.
30. Shoupe D, Mischell D Jr, Lacarra M, Lobo R, Horenstein J, d'Ablaing G. Correlation of endometrial maturation with four methods of estimation of day of ovulation. Obstet Gynecol 1989;73:88–92.
31. Lenton EA, Landgren BM, Sexton L. Normal variation in the length of the luteal phase of the menstrual cycle: identification of the short luteal phase. Br J Obstet Gynaecol 1984;91:685–9.
32. Smith SK, Lenton EA, Landgren BM, Cooke ID. The short luteal phase and infertility. Br J Obstet Gynaecol 1984;91:1120–2.
33. Ross GT, Cargille CM, Lipsett MB, Rayford PL, Marshall JR, Strott CA, et al. Pituitary and gonadal hormones in women during spontaneous and induced ovulatory cycles. Rec Prog Horm Res 1970;26:1–62.

FERTILITY AND STERILITY

Vol. 63, No. 2, February 1995
Printed on acid-free paper in U. S. A.

Corpus Luteum Defect—"Alloyed Gold Standard"

To the Editor:

We read with interest the review by Jordan and colleagues (1) of the tests currently available for the diagnosis of luteal phase defect (LPD). We agree that LPD is a heterogeneous disorder, with controversy surrounding its existence, diagnosis, and treatment. Jordan et al. contend that because all currently available tests for the diagnosis of LPD in some way relate to the secretion or action of P, that daily measurement of plasma P is the reference test against which all others should be judged.

It is becoming clear that this assumption may not be true. Many different hormonal regimens can adequately prepare the endometrium for cryopreserved and donor embryos. Receptivity of the endometrium to P can vary, independent of serum P levels; histologic delay can be present with physiologic P (2) or despite supraphysiologic P (3). Recently Batista et al. (4) demonstrated that integrated serum P is not a good predictor of the endometrial histology. On this basis, we feel that a single statement about the adequacy of P secretion may not be possible. From the perspective of a blastocyst entering the uterine cavity, LPD is primarily an endometrial disorder. All current treatment strategies attempt to advance the endometrial histology. As better markers of uterine receptivity become available (5), it is likely that LPD will be viewed as a subset of uterine receptivity defects. Until that time, debate will continue about the most appropriate way to diagnose (and treat) LPD.

Arthur J. Castelbaum, M.D.
Northern Fertility & Reproductive Associates
Meadowbrook, Pennsylvania

Bruce A. Lessey, M.D., Ph.D.
Department of Obstetrics and Gynecology
University of North Carolina
Chapel Hill, North Carolina
July 28, 1994

REFERENCES

1. Jordan J, Craig K, Clifton DK, Soules MR. Luteal phase defect: the sensitivity and specificity of diagnostic methods in common clinical use. Fertil Steril 1994;62:54–62.
2. Massai M-R, de-Zeigler D, Lesobre V, Bergeron C, Frydman R, Bouchard P. Clomiphene citrate affects cervical mucus and endometrial morphology independently of the changes in plasma hormonal levels induced by multiple follicular recruitment. Fertil Steril 1993;59:1179–86.
3. Toner JP, Hassiakos DK, Muasher SJ, Hsiu JG, Jones HW. Endometrial receptivities after leuprolide suppression and gonadotropin stimulation: histology, steroid receptor concentrations, and implantation rates. Ann NY Acad Sci 1991;622:220–9.
4. Bastista MC, Cartledge TP, Merino MJ, Axiotis C, Platia MP, Merriam GR, et al. Midluteal phase endometrial biopsy does not accurately predict luteal function. Fertil Steril 1993;59:294–300.
5. Lessey BA, Castelbaum AJ, Buck CA, Ying L, Yowell CW, Sun J. Further characterization of endometrial integrins during the menstrual cycle and in pregnancy. Fertil Steril 1994;62:497–506.

Reply of the Authors:

We appreciate the opportunity to reply to the letter of Drs. Castelbaum and Lessey. We share many areas of agreement in regards to luteal phase deficiency (LPD). While P is the primary known inducer of endometrial maturation, we are certain that medical science has only begun to identify all the components and factors that actually lead to an integrated mature endometrium that is the proper environment for implantation.

Our study was primarily designed to look at tests that are currently used or have been used for the diagnosis of LPD in the clinical setting (1). These tests include basal body temperature charts, endometrial biopsy, serum P levels in various combinations, and ovarian follicular ultrasound measurements. The temperature charts and serum P levels are both estimates and/or indicators of P secretion, while endometrial histology is directly influenced by P secretion. Since we can estimate total P secretion over the luteal phase in a fairly accurate manner (integrated daily serum P), we used this as our standard for comparison for the clinical tests. We agree that any test that estimates the adequacy of P secretion, be it integrated total P over the luteal phase or a single P level, is a step or more removed from the target tissue of interest—the endometrium. However, looking at endometrial histology itself, such as the interpretation of an endometrial biopsy, can only be a rough indicator of true biochemical maturation. Unfortunately, over the years, there have been no identified endometrial

markers that accurately reflect endometrial maturation and receptivity to implantation. Previous studies and investigators have looked at both E_2 and P receptors in the endometrium, and protein products of the endometrium that are found in the circulation (plasma endometrial protein [PEP], also known as PP14). The studies of endometrial steroid receptors have been contradictory and the PEP studies did not correlate with other indicators of endometrial maturation (2, 3). Although we would agree with Drs. Castelbaum and Lessey that a good endometrial marker for maturation and implantation would be attractive, such a marker has eluded previous investigations. We would predict such endometrial markers would be strongly influenced by prevailing serum and tissue levels of P. If endometrial markers of uterine receptivity become available, these certainly could become the ideal tests for corpus luteum function. We would predict that it will be difficult to establish an endometrial marker as the standard for the diagnosis of LPD because it will be as difficult to find a "gold standard" for comparison purposes as it has been in performing studies based on P levels. A true gold standard for corpus luteum and endometrial function would rely on specific findings that are present in a spontaneous successful conception cycle. However, the pregnancy (e.g., hCG) will likely alter the marker itself and change the findings that otherwise would be present in a normal nonconceptive luteal phase. The reference point for the endometrial marker may thus remain elusive and the investigators may have to rely on a secondary or indirect standard. In summary, we basically agree with the comments that were offered and would encourage these authors in their studies toward finding an endometrial marker for implantation and improve our diagnostic capabilities for LPD.

Michael R. Soules, M.D.
Donald K. Clifton, Ph.D.
Kristin Craig, B.S.
Division of Reproductive Endocrinology and Infertility
Department of Obstetrics and Gynecology
University of Washington School of Medicine
Seattle, Washington

John Jordan, M.D.
Everett, Washington
August 30, 1994

REFERENCES

1. Jordan J, Craig K, Clifton DK, Soules MR. Luteal phase defect: the sensitivity and specificity of diagnostic methods in common clinical use. Fertil Steril 1994;62:54–62.
2. Batista MC, Bravo N, Cartledge TP, Loriaux DL, Merriam GR, Neiman LK. Serum levels of placental protein 14 do not accurately reflect histologic maturation of the endometrium. Obstet Gynec 1993;81:439–43.
3. McNeely MJ, Soules MR. The diagnosis of luteal phase deficiency: a critical review. Fertil Steril 1988;50:1–15.

Diagnosis and Origins of Endometriomas

To the Editor:

We read with interest the recent article by Brosens et al. (1). We congratulate the authors on their new technique. However, in referring to our publication regarding endometriomas (2), the authors do not mention the difference between the invagination theory and our theory of the origin of ovarian endometriomas.

According to our paper, ovarian endometriomas are either primary or secondary. Primary (I) endometriomas are usually less than 3 cm in size, difficult to remove from the ovary due to fibrosis caused by endometriosis, and show endometriosis on histological examination. The origin of these endometriomas is similar to endometriosis in extra-ovarian sites in that the size of the cyst is limited by fibrosis and scarring.

Secondary (II) "endometriomas" are usually greater than 3 cm in size (3–20 cm in our series) and are not considered "true endometriomas." These "endometriomas" were originally functional (luteal or follicular) cysts that have been invaded by superficial ovarian cortex endometriosis. These "endometriomas" are subdivided into three groups based on their clinical appearance: IIA: cortical endometriosis plaques have not invaded the cyst wall and are easily removed from the ovarian stroma; IIB: similar to type A except the cortical endometriosis plaques have invaded the cyst wall. They are easily removed from the ovarian stroma except the one area that invades the cyst wall; IIC, the cortical endometriosis has invaded the major portion of the cyst wall. The cysts usually are attached to the pelvic sidewall or posterior broad ligaments by adhesions and fibrosis. They will often rupture at the time of separation from the pelvic sidewall.

Histologic evaluation of type IIA, IIB, and IIC "endometriomas" showed endometriosis in 0%,