*Journal of Athletic Training* 2016;51(1):47–56
doi: 10.4085/1062-6050-51.3.01

www.natajournals.org

*original research*

# Evaluation of an Algorithm to Predict Menstrual-Cycle Phase at the Time of Injury

**Timothy W. Tourville, PhD, ATC, CSCS*; Sandra J. Shultz, PhD, ATC, FNATA, FACSM†; Pamela M. Vacek, PhD‡; Emily J. Knudsen, PA-C, ATC§; Ira M. Bernstein, MD||; Kelly J. Tourville, MEd, ATC*; Daniel M. Hardy, PhD¶; Robert J. Johnson, MD#; James R. Slauterbeck, MD#; Bruce D. Beynnon, PhD#**

*Department of Rehabilitation and Movement Science, University of Vermont, Burlington; †Department of Kinesiology, University of North Carolina at Greensboro; ‡Medical Biostatistics Unit, University of Vermont, Burlington; §Dana Farber Cancer Institute, Boston, MA; ||Department of Obstetrics, Gynecology and Reproductive Sciences, University of Vermont, Burlington; ¶Department of Cell Biology and Biochemistry, Texas Tech University, Lubbock; #Department of Orthopaedics and Rehabilitation, University of Vermont, Burlington

***Context:*** Women are 2 to 8 times more likely to sustain an anterior cruciate ligament (ACL) injury than men, and previous studies indicated an increased risk for injury during the preovulatory phase of the menstrual cycle (MC). However, investigations of risk rely on retrospective classification of MC phase, and no tools for this have been validated.

***Objective:*** To evaluate the accuracy of an algorithm for retrospectively classifying MC phase at the time of a mock injury based on MC history and salivary progesterone (P4) concentration.

***Design:*** Descriptive laboratory study.

***Setting:*** Research laboratory.

***Participants:*** Thirty-one healthy female collegiate athletes (age range, 18–24 years) provided serum or saliva (or both) samples at 8 visits over 1 complete MC.

***Main Outcome Measure(s):*** Self-reported MC information was obtained on a randomized date (1–45 days) after mock injury, which is the typical timeframe in which researchers have access to ACL-injured study participants. The MC phase was classified using the algorithm as applied in a stand-alone computational fashion and also by 4 clinical experts using the algorithm and additional subjective hormonal history information to help inform their decision. To assess algorithm accuracy, phase classifications were compared with the actual MC phase at the time of mock injury (ascertained using urinary luteinizing hormone tests and serial serum P4 samples). Clinical expert and computed classifications were compared using $\kappa$ statistics.

***Results:*** Fourteen participants (45%) experienced anovulatory cycles. The algorithm correctly classified MC phase for 23 participants (74%): 22 (76%) of 29 who were preovulatory/anovulatory and 1 (50%) of 2 who were postovulatory. Agreement between expert and algorithm classifications ranged from 80.6% ($\kappa = 0.50$) to 93% ($\kappa = 0.83$). Classifications based on same-day saliva sample and optimal P4 threshold were the same as those based on MC history alone (87.1% correct). Algorithm accuracy varied during the MC but at no time were both sensitivity and specificity levels acceptable.

***Conclusions:*** These findings raise concerns about the accuracy of previous retrospective MC-phase classification systems, particularly in a population with a high occurrence of anovulatory cycles.

***Key Words:*** anterior cruciate ligament injury, risk factors, validation

**Key Points**

- Neither the algorithm nor our clinical expert assessment was able to accurately predict menstrual-cycle (MC) phase at the time of injury. In particular, specificity of the postovulatory phase could not be adequately assessed due to the high number of anovulatory cycles observed.
- These findings raise substantial questions regarding the accuracy of retrospectively determining the MC phase of young athletes, with or without a hormone measurement, in prior investigations, particularly in a population with a high occurrence of anovulatory cycles and a large prevalence of luteal-phase defects.
- Accurate determination of MC phase may only be possible through a prospective examination that captures estradiol and progesterone concentrations over multiple days before and after injury, which may not be logistically feasible in injury or disease risk-factor investigations in this population.
- These findings are particularly important for any investigation designed to retrospectively characterize MC phase and its association with injury or disease in this population.

Rupture of the anterior cruciate ligament (ACL) produces substantial disability and greatly increases the risk of early-onset posttraumatic osteoarthritis. Consequently, much effort has focused on identifying risk factors that predispose individuals to ACL injury. Although the risk of ACL injury is likely multifactorial,[1,2] a substantial body of literature[3–12] suggests that risk may differ across phases of a female's menstrual cycle (MC).

Sex-hormone concentrations are known to vary widely across days of the MC, and previous investigations have shown how normal physiologic changes in hormone concentrations can affect collagen metabolism[13] and ligament behavior[14]; they may also influence knee laxity, but the evidence for this in the literature is conflicting.[14,15] Sex-hormone fluctuations may subsequently influence other suspected ACL injury risk factors in a cyclic manner (eg, serum relaxin levels,[16] musculotendinous stiffness,[17] and knee-valgus motion[15]).

Multiple investigators[3–12] have examined the relationship between MC phase and the risk of sustaining an ACL injury. The consensus of these authors is that the risk of ACL injury may be disproportionately higher during the preovulatory phase, with some noting a higher proportion of injuries near menses, whereas others noted a higher proportion near ovulation.[5,6,9,18] Because of their retrospective designs, most of these researchers characterized MC phase based on historical data about a female's MC combined with calendar-counting methods.[3,4,6–9,11] These retrospective methods are limited in their ability to accurately identify the hormonal milieu (and hence MC phase) at the time of injury due to inconsistencies in participant recall[12] and variability in MC characteristics (eg, timing of hormone changes), even among those with 28- to 32-day cycles.[19,20] These limitations are particularly concerning when the injury event occurs near menses or ovulation, when hormone values are changing rapidly. To address these limitations, the authors[5,10,12] of 3 studies evaluated hormone concentrations soon after injury (2 to 72 hours) and found a greater-than-expected proportion of injuries near menses,[10] ovulation,[12] and more generally in the preovulatory phase.[5] Although hormone measurements were thought to increase the accuracy of phase determinations,[5,12] the sources of hormone data differed in these 3 studies, and the methods used to classify MC phase were not validated. In addition, it may not always be feasible to obtain hormone samples so near the time of injury.

Despite their limitations, case-control studies may be the only practical approach for exploring relationships between MC phase and injury risk. However, we are not aware of any studies that have compared information obtained at or after injury with that obtained prospectively to assess the validity of this approach. To conduct such an investigation, we formulated an algorithm based on retrospectively acquired hormonal and menstrual-history data and salivary progesterone concentrations to approximate the methods used to classify MC phase in previous retrospective ACL-injury risk-factor studies. We then comprehensively evaluated the algorithm's ability to determine MC phase (preovulatory versus postovulatory) retrospectively at the time a mock injury (or event) occurred using prospectively acquired data to determine true menstrual phase on the date of mock injury. We also determined if different investigators, who may be influenced by their own clinical judgment, could reliably apply the algorithm. We expected that the algorithm would yield consistent results across investigators but that accurate classification of MC phase at the time of mock injury would be a challenge. Our evaluation therefore included an examination of how salivary progesterone (P4) and the timing of sample collection influence algorithm performance. We also examined whether the algorithm's accuracy varied depending on when during the MC the mock injury occurred.

## MENSTRUAL-PHASE ALGORITHM

An algorithm was developed to categorize participants as preovulatory or postovulatory at the time of injury using self-reported menstrual history and a salivary P4 concentration obtained after the injury date (Figure). This algorithm was based on published data[21–23] regarding MCs in young women, as well as clinical experience. Similar to many of the methods based on calendar counting, the algorithm assumes that the length of the luteal phase is more stable than that of the follicular phase, with ovulation usually occurring 14 days before the onset of menses. In accordance with published data,[22,24] a salivary P4 concentration of 190 pmol/L was selected as the threshold indicating that ovulation occurred. Participants with injuries occurring within 14 days of the next menses are considered *anovulatory* if their P4 values are below the 190 pmol/L threshold and their observed cycle is 3 or more days shorter or longer than their normally reported MC length.

## METHODS

Thirty-three National Collegiate Athletic Association Division I female athletes between the ages of 18 and 24 years from the University of Vermont were examined in this study. All participants reported a normal MC history, no use of prescription medications or illicit substances, no use of tobacco over the past 12 months, and good overall health with no underlying conditions. They also met a weekly physical activity energy expenditure of 60 kcal/kg/wk or greater during moderate-, high-, and very high-intensity activities.[25] *Normal MC history* was defined as $>$25 days and $<$40 days between cycles with no change in menstrual status in the past 12 months. Energy expenditure during physical activity was measured with the 7-Day Physical Activity Recall (PAR).[25] Our institutional review board approved this investigation, and all participants provided written informed consent before enrollment.

### Prospective Characterization of the MC

Participants visited the laboratory on 8 occasions, spaced equally over the course of 1 MC. The first day of testing began within 24 hours of the onset of menses, and spacing of subsequent visits was based on the length of her *previous MC,* which was defined as the number of days from the first day of bleeding during the previous MC to the last day before the onset of bleeding of the next MC. During visits 1, 3, 5, and 7, participants provided saliva and serum samples. At visits 2, 4, 6, and 8, subjects provided saliva samples and completed the PAR questionnaire.[25,26] Unless the next menses began before visit 8, its start date was obtained through telephone contact.

**Salivary Sample Collection and Analysis.** We instructed participants to abstain from consuming alcohol for 12 hours before sample collection and consuming anything by mouth (except water) or chewing gum for a minimum of 1 hour before each visit and to refrain from flossing their teeth between waking and their laboratory visit. These restrictions normalize salivary pH and avoid potential blood contamination. A minimum of 10 minutes

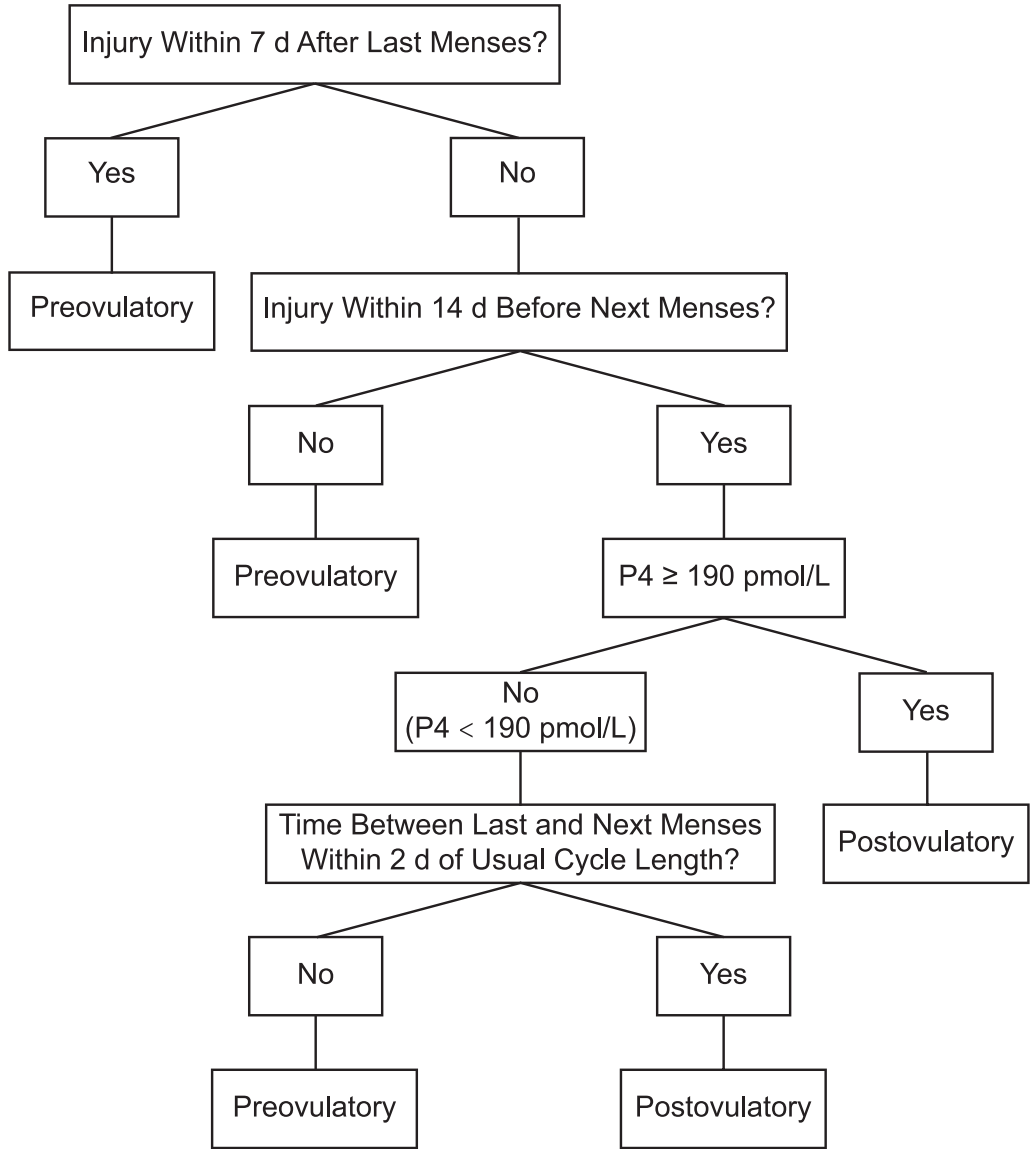


**Figure. Mathematical menstrual-cycle phase classification algorithm. Abbreviation: P4, Progesterone concentration.**

before collection, participants rinsed their mouths with water 3 times for 10 seconds per rinse to remove food particulates and normalize salivary pH.[27,28] Approximately 4 mL of whole, nonstimulated saliva was collected by passive drool through a 5-cm long plastic straw. Samples were stored at –80°C until shipment to the Oregon National Primate Research Center (ONPRC) at Oregon Health Sciences University for analysis using Salimetrics Salivary Progesterone immunoassay kits (Salimetrics LLC, State College, PA). The intra-assay and interassay coefficients of variation (CVs) for this assay as performed by ONPRC were 5.63% and 8.37%, respectively.

**Serum Sample Collection and Analysis.** We obtained approximately 8-mL samples of whole blood via standard venipuncture and stored them at –80°C until they were shipped to ONPRC for analysis. Serum P4 was analyzed in duplicate on an Immulite 2000 (Siemens Healthcare, Munich, Germany) automated clinical immunoassay system. Intra-assay and interassay CVs were less than 10%.

**Assessment of Day of Ovulation.** Starting on day 8 of their MCs, participants were instructed to perform non–first-morning void luteinizing hormone-based ovulation tests (Kurkel Enterprises, LLC, Redmond, WA) to detect the rise in urinary luteinizing hormone that indicates ovulation date. The tests were to be used at the same time each day, between 11:00 AM and 8:00 PM as recommended by the manufacturer. Each participant underwent 2 educational sessions regarding ovulation test procedures and result documentation before beginning the tests. Results were recorded on a log sheet and discussed with investigators at each visit to ensure accurate documentation and compliance. Tests were either stopped after a positive result or continued until the first day of menstruation during the subsequent cycle. If a test was considered invalid (ie, inconclusive or suspected to be faulty), we instructed participants to repeat it at the next normal void.

**Assignment of Mock-Injury Date.** At the time of enrollment, participants were randomly assigned a mock-

**Table 1. Participant Demographics (N = 31)**

| Characteristic | Mean ± SD | Range, Minimum–Maximum |
|---|---|---|
| Age, y | 19.80 ± 1.30 | 18.00–23.00 |
| Height, cm | 167.40 ± 7.90 | 152.40–185.40 |
| Mass, kg | 66.40 ± 9.80 | 55.70–98.90 |
| Body mass index | 23.50 ± 2.60 | 19.50–31.70 |
| Waist-to-hip ratio | 0.74 ± 0.03 | 0.69–0.79 |
| Body fat, % | 23.80 ± 5.90 | 14.30–40.10 |

**Table 2. Salivary Progesterone Concentrations[a]**

| | | Concentration, pmol/L | |
|---|---|---|---|
| Sample | No. | Mean ± SD | Range |
| From anovulatory women | 110 | 255 ± 154 | 62–1110 |
| During preovulatory phase | 82 | 230 ± 99 | 49–447 |
| During postovulatory phase | 42 | 439 ± 253 | 82–1414 |

[a] All samples were collected during the same menstrual cycle.

injury event date that would be subsequently used to evaluate the algorithm's accuracy in identifying the MC phase at time of injury. In an effort to simulate the recall of the date of significant musculoskeletal trauma, such as an ACL tear, participants were not informed of the mock injury until the scheduled date, when they were given a brightly colored bracelet displaying the date in large, bold text. They were not required to wear the bracelet; however, they were asked to place it in a location they would observe multiple times throughout the day. At each subsequent visit, we stressed the importance of this mock-injury date. The motivation for providing a date-inscribed item and stressing the importance of remembering the date was to create a memorable event date that would serve as a surrogate for the point in time when an athlete actually sustained a significant injury.

### Retrospective Classification of MC Phase

To assess the accuracy of the algorithm to retrospectively determine MC phase at the time of mock injury, participants were interviewed by an investigator (K.J.T.) who was blinded to all aspects of the study relating to MC characterization. She acquired MC information using a modified version of the Hormonal History Questionnaire (HHQ), which has been previously validated and applied.[12] The HHQ information included average MC length, normal days of menstruation, date of the first day of the last menstrual period, average number of days between periods, number of menstrual periods in the past 12 months, premenstrual symptoms and their severity, use of hormonal therapies of any kind, and the date of the mock injury. Random numbers (1–45) were assigned to specify the number of days after the mock injury that attempts to contact the participant and administer the HHQ could begin. This was done to replicate, as best as possible, real-world variations in the time between injury and enrollment in a retrospective case-control investigation, because the recall accuracy of HHQ information for a particular MC declines as time progresses.[29] Similarly, in most studies of ACL injury risk, it is not always possible to obtain a saliva sample on the day of injury. We therefore randomly selected a salivary sample from among those obtained on or after the mock-injury date but before the start of the next menses.

The algorithm was applied to these retrospectively acquired data using 2 methods to classify MC phase at the time of mock injury. First, it was applied in a purely objective fashion using a sequence of *yes*/*no* responses (Figure). Second, 4 investigators (S.J.S., I.M.B., D.M.H., J.R.S.) applied the algorithm using the same salivary P4 concentrations and self-reported menses dates. The investigators had backgrounds typical of researchers interested in studying the relationship between MC phase and injury risk.

### Algorithm Validation

To determine the validity of the algorithm when applied purely objectively and then by each of the 4 investigators, each woman's assignment into the preovulatory or postovulatory phase of the MC at the time of the mock injury was compared with her known phase on that date as determined by an expert in reproductive medicine (I.M.B.) using the prospectively acquired data (luteinizing hormone-based ovulation detection kits, serial serum P4 concentrations, and menses dates). We conducted secondary analyses to examine whether algorithm accuracy could be improved by using salivary P4 samples obtained on the mock-injury date or by modifying the algorithm to include MC-history information only. To determine how the timing of injury during an MC influences algorithm performance, we examined its accuracy when each of the 8 visits was selected as the mock-injury date. Salivary P4 concentrations were analyzed using mixed-model linear regression to test the differences between MC phases and estimate within-person and between-persons variability.

The following accuracy measures were computed. *Sensitivity* was the proportion of women in the preovulatory phase who were correctly classified as preovulatory. *Specificity* was the proportion of women in the postovulatory phase who were correctly classified as postovulatory. *Positive predictive value (PPV)* was the proportion of women classified as preovulatory who were actually in the preovulatory phase. *Negative predictive value (NPV)* was the proportion of women classified as postovulatory who were actually in the postovulatory phase. *Overall accuracy* was the proportion of all women who were correctly classified.

We used κ statistics to assess agreement in phase assignments between investigators and the mathematical algorithm. Pearson product moment correlation was used to examine the relationship between salivary and serum P4 concentrations. All data analyses were performed using SAS (version 9.2; SAS Institute Inc, Cary, NC).

## RESULTS

Two athletes were excluded during the course of the study: 1 participant did not complete all data-collection requirements, and 1 had an average PAR <60 kcal/kg/wk during the 4-week testing interval. Demographics for the remaining 31 women are listed in Table 1; their average PAR score was 78.9 ± 30.2 kcal/kg/wk. Overall, the mean salivary P4 concentration was higher for samples obtained during the postovulatory phase (as determined based on prospectively acquired serum samples and ovulation tests) than for samples obtained during the preovulatory phase

and from anovulatory women (Table 2). However, the estimated within-phase variability for samples taken from the same woman (CV = 0.50) was as large as the variability between phases (CV = 0.49) and much larger than the variability among women (CV = 0.31). The overall correlation between salivary and serum P4 concentrations taken on the same day was $R = 0.71$.

Menstrual-cycle length ranged from 19 to 60 days (mean = 31 ± 9.5 days). The mock-injury date was, on average, 10.7 days after the onset of menses (range = 0–31 days). The average number of days to the next menses after the injury date was 19.8 days. Based on the prospectively acquired data, 14 of the 31 participants (45.2%) were anovulatory during the tested cycle, whereas 15 (48.4%) were preovulatory and 2 (6.5%) were postovulatory at the time of their mock injury. The HHQs were administered an average of 20 days after the mock-injury date (range = 6–54 days), and 23 athletes (74%) accurately recalled the starting dates of the menses before and after the mock injury. Six participants reported incorrect dates for the menses before the mock injury, with differences ranging from −3 to +4 days, and 4 women (including 2 with incorrect starting dates) incorrectly reported the date of menses after the mock injury, with differences ranging from −3 to +5 days. The number of days between the mock injury and the randomly selected saliva sample used to evaluate the algorithm ranged from 0 to 35 days (median = 10 days).

Each athlete's self-reported MC data and salivary P4 concentration from the sample selected for use in the algorithm are shown in Table 3, along with the menstrual-phase classification at the time of the mock injury as determined by the algorithm, and her actual phase. The sensitivity of the algorithm to correctly classify the 29 participants in the preovulatory phase at the time of mock injury was 75.9%, whereas its specificity for correctly classifying the 2 postovulatory participants was 50% (Table 4). The PPV of a preovulatory/anovulatory classification by the algorithm was high (95.6%), due to the high prevalence of this ovulatory phase (93.5%), which corresponds to the PPV that would be achieved using a completely random phase assignment. The NPV of a postovulatory classification was 12.5%.

One postovulatory woman was incorrectly classified by the algorithm as preovulatory/anovulatory because her mock injury occurred after ovulation but more than 14 days before her next menses. Of the 8 athletes that the algorithm classified as postovulatory at the time of injury, only 1 was actually postovulatory, 5 were anovulatory, and 2 were preovulatory, as determined prospectively based on serial serum P4 concentrations and ovulation tests (Table 3). The 2 preovulatory women (participants 28 and 30) had mock injuries within 14 days of the next menses but ovulated after the mock-injury date. They both had salivary P4 concentrations well above 190 pmol/L during their entire cycles, so the timing of their saliva samples did not contribute to the misclassification. Of the 5 anovulatory individuals, 2 (participants 26 and 29) had MCs that were shorter than their average cycles but were classified by the algorithm as postovulatory because of their high salivary P4 concentrations. Conversely, participants 24 and 25 were misclassified as postovulatory despite their low salivary P4 concentrations because their MCs were the usual length. Participant 27 was misclassified because on the HHQ she erroneously recalled her menses date as being 13 days after the mock-injury date, though the actual date was 16 days after injury.

The menstrual-phase classifications obtained when the 4 investigators applied the algorithm differed somewhat from the objectively computed classifications, with agreement ranging from 80.6% ($\kappa = 0.50$, 95% confidence interval = 0.15, 0.84) to 93.5% ($\kappa = 0.83$, 95% confidence interval = 0.60, 1.00). Sensitivity ranged from 72.4% to 75.9%, which was comparable with the mathematically applied algorithm. Three investigators correctly classified 1 of the 2 postovulatory women (50% specificity), and the fourth correctly classified both (100% specificity).

When the objectively computed algorithm was based on saliva samples taken on the date of mock injury, rather than the randomly selected sample, accuracy showed little improvement (Table 4). The algorithm correctly classified 2 preovulatory/anovulatory athletes who were previously misclassified and misclassified 1 preovulatory/anovulatory athlete who had previously been classified correctly. Modifying the algorithm so that it was based only on the self-reported MC data improved sensitivity to 89%. It also improved both PPV and NPV, but NPV and specificity both remained very low (Table 4).

The accuracy of the algorithm varied widely depending on timing of the mock injury (Table 5). Sensitivity was highest when the mock injury occurred early in the MC because all or most participants were in the preovulatory phase, but it dropped steeply at midcycle, when the algorithm classified many preovulatory participants as postovulatory. In contrast, specificity and NPV were higher when the mock injury occurred later in the MC. There was no time during the MC when both the sensitivity and specificity of the algorithm were acceptable.

## DISCUSSION

Because ACL injuries occur infrequently, case-control studies remain the most practical research design to explore relationships between MC phase and ACL injury risk. For these retrospective studies to be valid, it is necessary to establish an objective procedure (algorithm) to accurately classify an individual's MC phase at the time of injury based on data that can be reasonably obtained after injury. Consequently, the goal of our study was not to create a novel algorithm but rather to evaluate our ability to classify phase using an algorithm based on current MC-phase knowledge, as has been used by previous investigators attempting to classify phase after injury. Our evaluation of an objective algorithm based on a menstrual-history questionnaire and single salivary P4 sample obtained after a mock injury indicates that these data were insufficient for accurate retrospective classification of participants as preovulatory/anovulatory versus postovulatory at the time of a mock injury.

When applied objectively, the algorithm correctly classified only 74% of the 31 participants as being preovulatory versus postovulatory at the time of injury. Similar results were obtained when experts in the field applied the algorithm, reflecting their 80.6% to 93.5% agreement with the objectively applied algorithm. Although this suggests that the algorithm could yield similar results from different investigators, it was unable to classify

**Table 3. Algorithm Input and Outcome Data: Comparison of Predicted and Actual Menstrual Phase on the Day of Mock Injury**

| Participant | Self-Reported Menstrual-Cycle Information | | | | Salivary Progesterone Concentration, pmol/L | Injury to Saliva Sample, d | Phase Classification | Actual Phase[a] | Incorrect Prediction |
|---|---|---|---|---|---|---|---|---|---|
| | Menses Onset to Injury, d | Injury to Next Menses Onset, d | Days in Cycle | Average Days in Cycle | | | | | |
| 1 | 4 | 21 | 25 | 28.0 | 59 | 21 | Preovulatory/anovulatory | Preovulatory | |
| 2 | 0 | 32 | 32 | 29.0 | 94 | 16 | Preovulatory/anovulatory | Preovulatory | |
| 3 | 12 | 16 | 28 | 26.5 | 142 | 4 | Preovulatory/anovulatory | Preovulatory | |
| 4 | 31 | 15 | 46 | 37.5 | 144 | 0 | Preovulatory/anovulatory | Postovulatory | X |
| 5 | 13 | 7 | 20 | 24.0 | 153 | 3 | Preovulatory/anovulatory | Anovulatory | |
| 6 | 8 | 15 | 23 | 28.0 | 179 | 14 | Preovulatory/anovulatory | Anovulatory | |
| 7 | 4 | 31 | 35 | 30.0 | 188 | 18 | Preovulatory/anovulatory | Preovulatory | |
| 8 | 9 | 19 | 28 | 28.0 | 188 | 8 | Preovulatory/anovulatory | Anovulatory | |
| 9 | 8 | 20 | 28 | 31.0 | 199 | 14 | Preovulatory/anovulatory | Preovulatory | |
| 10 | 15 | 45 | 60 | 37.5 | 215 | 21 | Preovulatory/anovulatory | Anovulatory | |
| 11 | 8 | 22 | 30 | 29.0 | 221 | 13 | Preovulatory/anovulatory | Preovulatory | |
| 12 | 6 | 21 | 27 | 28.0 | 244 | 0 | Preovulatory/anovulatory | Anovulatory | |
| 13 | 1 | 30 | 31 | 30.0 | 259 | 15 | Preovulatory/anovulatory | Preovulatory | |
| 14 | 1 | 30 | 31 | 28.5 | 262 | 7 | Preovulatory/anovulatory | Preovulatory | |
| 15 | 7 | 23 | 30 | 28.0 | 334 | 0 | Preovulatory/anovulatory | Anovulatory | |
| 16 | 3 | 25 | 28 | 28.0 | 341 | 10 | Preovulatory/anovulatory | Preovulatory | |
| 17 | 8 | 21 | 29 | 28.0 | 342 | 20 | Preovulatory/anovulatory | Preovulatory | |
| 18 | 1 | 29 | 30 | 29.0 | 361 | 17 | Preovulatory/anovulatory | Preovulatory | |
| 19 | 10 | 20 | 30 | 34.0 | 385 | 0 | Preovulatory/anovulatory | Anovulatory | |
| 20 | 13 | 17 | 30 | 30.0 | 483 | 8 | Preovulatory/anovulatory | Preovulatory | |
| 21 | 0 | 28 | 28 | 28.0 | 638 | 0 | Preovulatory/anovulatory | Anovulatory | |
| 22 | 8 | 19 | 27 | 28.0 | 709 | 14 | Preovulatory/anovulatory | Preovulatory | |
| 23 | 0 | 40 | 40 | 32.0 | 1111 | 35 | Preovulatory/anovulatory | Anovulatory | |
| 24 | 23 | 9 | 32 | 31.5 | 85 | 9 | Postovulatory | Anovulatory | X |
| 25 | 13 | 13 | 26 | 25.0 | 158 | 0 | Postovulatory | Anovulatory | X |
| 26 | 16 | 12 | 28 | 32.0 | 280 | 12 | Postovulatory | Anovulatory | X |
| 27 | 8 | 13 | 21 | 28.0 | 330 | 0 | Postovulatory | Anovulatory | X |
| 28 | 18 | 13 | 31 | 28.0 | 370 | 4 | Postovulatory | Preovulatory | X |
| 29 | 9 | 14 | 23 | 28.0 | 418 | 0 | Postovulatory | Anovulatory | X |
| 30 | 14 | 13 | 27 | 29.0 | 923 | 14 | Postovulatory | Preovulatory | X |
| 31 | 23 | 5 | 28 | 28.5 | 394 | 10 | Postovulatory | Postovulatory | |

[a] Actual cycle phase on mock-injury date as determined prospectively using serial serum progesterone samples and ovulation tests.

**Table 4. Accuracy of the Menstrual-Phase Algorithm, %**

| | Original Algorithm | | Modified Algorithm (Without Salivary Progesterone) |
|---|---|---|---|
| | Random Saliva Sample[a] | Saliva Sample on Injury Date | |
| Sensitivity for detecting a preovulatory/anovulatory phase | 75.9 | 79.3 | 89.7 |
| Specificity for detecting a postovulatory phase | 50.0 | 50.0 | 50.0 |
| Positive predictive value (preovulatory/anovulatory prediction) | 95.6 | 95.8 | 96.3 |
| Negative predictive value (postovulatory prediction) | 12.5 | 14.3 | 25.0 |
| Correctly classified | 74.2 | 77.4 | 87.1 |

[a] Saliva sample was selected randomly from those collected on or after the mock-injury date and before the next menses (results corresponding to data in Table 2).

participants with a high degree of accuracy. This lack of accuracy may be due to 3 primary sources of error: (1) the inability of athletes to correctly recall their menstrual history accurately, (2) assumption of a stable, constant, 14-day luteal-phase window, and (3) the inability of a single hormone sample to adequately capture the large variability in MC characteristics and hormone profiles across women.

The study was designed to initiate contact with study participants for the purpose of acquiring menstrual-history data at a random time interval (1–45 days) by a blinded investigator after the mock-injury date to mimic real-world situations in which it may not always be feasible to obtain data immediately after the injury, particularly in large-scale studies. We were able to make contact with the participants and acquire their menstrual-history data between 6 and 54 days after the mock injury. These data were then used to identify the start date for the menses occurring before and after the mock-injury date. Studies accessing self-reported MC information indicate there may be substantial recall error, particularly in younger females who have more variable cycle lengths.[30] Moreover, recall error may be particularly problematic if injury occurs near midcycle (ie, ovulation), when even a 1-day or 2-day error in menses dates could result in a different phase classification. In our study, 23 women (74.2%) correctly recalled the starting dates of the menses before and after the date of mock injury. Of the 8 women who erroneously reported 1 date or both dates, only 1 was misclassified because of the error, and her mock injury occurred midcycle, so recall error was not a major reason for the algorithm's poor performance.

In addition to its potential recall error, self-reported menstrual-history information cannot reveal the length of specific MC phases or the timing of ovulation. Although the average MC length is 28 days and ovulation on average occurs around day 14, the actual length of an individual cycle, the timing of ovulation, and thus the length of the follicular and luteal phases are all known to vary.[20,21,31–34] Specific to our methods, counting back 14 days from the start of the next cycle was based on the assumption that the length of the luteal phase is more consistent than that of the follicular phase.[5,9,17,31] Whereas the follicular phase is reported to average 13 to 14 days in length,[21,31,33] actual luteal-phase lengths range from 11 to 18 days in young females aged 18 to 24 years[21] (8–17 in females aged 18–40 years[31,33]). Moreover, some findings suggest that exercising women have significantly shorter luteal phases (eg, 8.2 $\pm$ 0.5 days).[35] In a study of 73 physically active females aged 18 to 30 years,[20] only 32% had a positive urinary ovulation test and 59% had attained a P4 level criterion of $>$2.0 ng/mL when counting back 12 to 14 days from the start of the next cycle. Hence, using a standard criterion of 14 days to represent the luteal-phase length in physically active women may contribute to the inaccuracy of the algorithm. In our study, the phases of 2 participants were misclassified because ovulation occurred less than 14 days before the start of the next menses. A summary of previous studies

**Table 5. Effect of Injury Timing on Algorithm Performance**

| | Visit Used as Mock-Injury Date[a] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Actual phase | | | | | | | | |
| No. preovulatory/anovulatory | 29 | 31 | 31 | 30 | 27 | 16 | 11 | 8 |
| No. postovulatory | 0 | 0 | 0 | 1 | 4 | 14 | 13 | 12 |
| Predicted phase[b] | | | | | | | | |
| No. preovulatory/anovulatory | 29 | 31 | 30 | 25 | 15 | 8 | 6 | 5 |
| No. postovulatory | 0 | 0 | 1 | 6 | 16 | 22 | 18 | 15 |
| Sensitivity for detecting a preovulatory/anovulatory phase, % | 100.0 | 100.0 | 96.8 | 80.0 | 51.9 | 37.5 | 54.5 | 50.0 |
| Specificity for detecting a postovulatory phase, % | [c] | [c] | [c] | 0.0 | 75.0 | 85.7 | 100.0 | 91.7 |
| Positive predictive value (preovulatory/anovulatory prediction), % | 100.0 | 100.0 | 100.0 | 96.0 | 93.3 | 75.0 | 100.0 | 80.0 |
| Negative predictive value (anovulatory prediction), % | [c] | [c] | 0.0 | 0.0 | 18.6 | 54.5 | 72.2 | 73.3 |
| Correctly classified, % | 100.0 | 100.0 | 96.8 | 77.4 | 54.8 | 60.0 | 79.2 | 75.0 |

[a] Saliva sample from the same visit was used to maximize algorithm performance.
[b] Visit 1 data exclude 2 participants who erroneously reported the date of last menses as after the first visit. Data for visits 6 to 8 exclude participants whose next menstrual phase began before these visits.
[c] Specificity could not be calculated because no participant was in the postovulatory phase at this test time.

**Table 6. Summary of Menstrual-Cycle Phase Classification Methods for Studies of the Relationship Between Phase of the Menstrual Cycle and Risk of Sustaining an Anterior Cruciate Ligament Injury**

| Study | Menstrual-Cycle Phase Assignment Method[a] | Anterior Cruciate Ligament Injuries, % | |
|---|---|---|---|
| | | Preovulatory Phase[b] | Postovulatory Phase[b] |
| Myklebust et al,[7] 1998 (N = 17) | Calendar counting | 41 | 59 |
| Wojtys et al,[11] 1998 (N = 28) | Calendar counting | 42 | 58 |
| Arendt et al,[4] 2002 (N = 83; n = 58 not taking oral contraceptives) | Calendar counting | Higher risk[c] | Lower risk[c] |
| Wojtys et al,[12] 2002 (N = 51) | Calendar counting and hormone assessment | 66 | 34 |
| Slauterbeck et al,[10] 2002 (N = 37) | Calendar counting and hormone assessment | 74 | 26 |
| Myklebust et al,[8] 2003 (N = 46) | Calendar counting | 76 | 24 |
| Beynnon et al,[5] 2006 (N = 46) | Calendar counting and hormone assessment | 74 | 26 |
| Adachi et al,[3] 2008 (N = 18) | Calendar counting | 83 | 17 |

[a] Calendar counting classification method implies menstrual-cycle questionnaire-based assessments.
[b] *Preovulatory phase* was defined as days 1 to 14 in a normal 28-day menstrual cycle. *Postovulatory phase* was defined as days 15 to 28 in a normal menstrual cycle.
[c] No specific percentage breakdown for anterior cruciate ligament injury was provided.

examining the relationship between MC phase and ACL injury risk is provided in Table 6.[3,5,7,8,10]

Given the challenges of these calendar-based methods, more recent investigations of the relationship between ACL injury risk and MC phase have evaluated hormonal concentrations near the time of injury (<72 hours) to better estimate the actual hormone milieu.[5,10,12] However, despite the inclusion of salivary P4 concentration in our algorithm, we were unable to achieve an acceptable level of accuracy in MC-phase determination, even when we used the sample obtained on the mock-injury date. We suspect this may be due to the choice of a 190-pmol/L threshold, as P4 concentrations observed in our study were higher than those reported by Chatterton et al.[22] Yet when we used the optimal threshold for classification in our study (340 pmol/L) and the saliva sample obtained on the day of injury, sensitivity and specificity were no better than when the algorithm was based on self-reported MC data alone.

Use of a single sample to determine MC phase has recently been questioned.[19] Because of the inherent variability in MC characteristics, it may be difficult to determine from a single sample whether hormone levels are rising, peaking, or falling. Hence, a single sample coupled with the known inaccuracies in calendar-based counting methods already discussed may not provide sufficient information to accurately identify the phase of the MC at the time of injury. Although research suggests that taking multiple samples around the event of interest may provide a better representation of the hormonal milieu than a single sample,[34] further study is needed to determine if this would improve the determination of MC phase. We did not explore the use of multiple salivary P4 samples to determine phase in our study because of their large variability. Samples obtained from the same woman differed as much within each MC phase as between phases, so she could have a preovulatory sample with a higher P4 concentration than a postovulatory sample.

We have 2 concerns regarding the high number of anovulatory cycles that occur in young athletic females, despite their normally occurring menses. First, we have concern that an altered hormonal milieu may affect their bone health, and this requires further attention. Second, with our algorithm, anovulatory participants meeting the third criterion of the algorithm (sample below salivary P4 threshold, evidence of menses within 14 days of injury, and onset of menses within 2 days of the normal MC length reported in the questionnaire) were incorrectly classified as postovulatory. This is particularly problematic for competitive athletes, as our reported frequency of anovulatory MCs (45%) was more than twice that for females exercising >2 h/wk at 55% of maximum heart rate[23] or between 2 and 10 hours for 3 months or longer.[34] It is possible that the algorithm could be modified to better distinguish between women in the postovulatory phase and those experiencing an anovulatory cycle. However, correct classification of women with anovulatory MCs that are of similar length to their usual MCs is likely to remain a problem unless more predictive hormone data are available.

Compounding these difficulties is the fact that errors in menstrual-phase assessment are more likely when injuries occur at certain times during the MC than others. We demonstrated this by examining how the sensitivity and specificity of our algorithm changed with the timing of the mock injuries. The accuracy of other methods of phase determination is also likely to vary depending on the timing of injury, which is of concern because it indicates that the inaccuracy reflects bias as well as imprecision. Although this was not the primary goal of our study, future investigators should evaluate the temporal response of sex hormones across multiple MCs, as well as the effects of physical and emotional stress on sex-hormone levels in collegiate-aged competitive athletes, as this information may aid in the ability to retrospectively classify MC phase after musculoskeletal trauma.

Whereas this study was limited to a relatively small sample of collegiate athletes with a high percentage of anovulatory cycles, they do represent the high-risk populations typically included in ACL injury risk-factor studies. Only 2 of the women were postovulatory at the time of mock injury, limiting our ability to assess the specificity with which the algorithm classified these individuals. However, the analysis discussed earlier, in which alternative mock-injury dates were examined, indicated that postovulatory women were often incorrectly

classified unless the injury occurred near the end of the MC.

## CONCLUSIONS

We showed that an algorithm using a single measurement of salivary P4 concentration to retrospectively classify MC phase was not valid for use in risk-factor studies. These findings raise substantial questions regarding the accuracy of prior investigations that retrospectively determined the MC phases of young athletes based on calendar methods with or without a single hormone sample, particularly in a population with such a high occurrence of anovulatory cycles.[19] It is possible that accuracy may be improved by acquiring multiple samples immediately after the day of injury, but our data indicate that salivary P4 may not be adequate for this purpose because of within-phase variability. Ultimately, accurate determination of MC phase may only be possible through prospective examination that captures both estradiol and progesterone concentrations over multiple days around (ie, both before and after) the time of injury. This would be virtually impossible for studies of risk for ACL trauma because a very large number of athletes would need to be tested over extended periods of time to accrue an adequate sample of injured participants.

## ACKNOWLEDGMENTS

Funding for this investigation was provided by research grant R01-AR050421 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health (Dr Beynnon).

We thank Hellen S. Hollenbach, PA-C, MS, ATC, and Sarah L. Davidson, MD, for their assistance with data collection. We also thank our research participants for their time and efforts with this investigation.

## REFERENCES

1. Renstrom P, Ljungqvist A, Arendt E, et al. Non-contact ACL injuries in female athletes: an International Olympic Committee current concepts statement. *Br J Sports Med*. 2008;42(6):394–412.
2. Shultz SJ, Schmitz RJ, Benjaminse A, Chaudhari AM, Collins M, Padua DA. ACL Research Retreat VI: an update on ACL injury risk and prevention. *J Athl Train*. 2012;47(5):591–603.
3. Adachi N, Nawata K, Maeta M, Kurozawa Y. Relationship of the menstrual cycle phase to anterior cruciate ligament injuries in teenaged female athletes. *Arch Orthop Trauma Surg*. 2008;128(5):473–478.
4. Arendt EA, Bershadsky B, Agel J. Periodicity of noncontact anterior cruciate ligament injuries during the menstrual cycle. *J Gend Specif Med*. 2002;5(2):19–26.
5. Beynnon BD, Johnson RJ, Braun S, et al. The relationship between menstrual cycle phase and anterior cruciate ligament injury: a case-control study of recreational alpine skiers. *Am J Sports Med*. 2006;34(5):757–764.
6. Lefevre N, Bohu Y, Klouche S, Lecocq J, Herman S. Anterior cruciate ligament tear during the menstrual cycle in female recreational skiers. *Orthop Traumatol Surg Res*. 2013;99(5):571–575.
7. Myklebust G, Maehlum S, Holm I, Bahr R. A prospective cohort study of anterior cruciate ligament injuries in elite Norwegian team handball. *Scand J Med Sci Sports*. 1998;8(3):149–153.
8. Myklebust G, Engebretsen L, Braekken IH, Skjolberg A, Olsen OE, Bahr R. Prevention of anterior cruciate ligament injuries in female team handball players: a prospective intervention study over three seasons. *Clin J Sport Med*. 2003;13(2):71–78.
9. Ruedl G, Ploner P, Linortner I, et al. Are oral contraceptive use and menstrual cycle phase related to anterior cruciate ligament injury risk in female recreational skiers? *Knee Surg Sports Traumatol Arthrosc*. 2009;17(9):1065–1069.
10. Slauterbeck JR, Fuzie SF, Smith MP, et al. The menstrual cycle, sex hormones, and anterior cruciate ligament injury. *J Athl Train*. 2002;37(3):275–278.
11. Wojtys EM, Huston LJ, Lindenfeld TN, Hewett TE, Greenfield ML. Association between the menstrual cycle and anterior cruciate ligament injuries in female athletes. *Am J Sports Med*. 1998;26(5):614–619.
12. Wojtys EM, Huston L, Boynton MD, Spindler KP, Lindenfeld TN. The effect of the menstrual cycle on anterior cruciate ligament injuries in women as determined by hormone levels. *Am J Sports Med*. 2002;30(2):182–188.
13. Shultz SJ, Wideman L, Montgomery MM, Beasley KN, Nindl BC. Changes in serum collagen markers, IGF-I, and knee joint laxity across the menstrual cycle. *J Orthop Res*. 2012;30(9):1405–1412.
14. Zazulak BT, Paterno M, Myer GD, Romani WA, Hewett TE. The effects of the menstrual cycle on anterior knee laxity: a systematic review. *Sports Med*. 2006;36(10):847–862.
15. Shultz SJ, Schmitz RJ, Kong Y, et al. Cyclic variations in multiplanar knee laxity influence landing biomechanics. *Med Sci Sports Exerc*. 2012;44(5):900–909.
16. Dragoo JL, Castillo TN, Korotkova TA, Kennedy AC, Kim HJ, Stewart DR. Trends in serum relaxin concentration among elite collegiate female athletes. *Int J Womens Health*. 2011;3:19–24.
17. Eiling W, Bryant AL, Petersen W, Murphy A, Hohmann E. Effects of menstrual-cycle hormone fluctuations on musculotendinous stiffness and knee joint laxity. *Knee Surg Sports Traumatol Arthrosc*. 2007;15(2):126–132.
18. Hewett TE, Zazulak BT, Myer GD. Effects of the menstrual cycle on anterior cruciate ligament injury risk: a systematic review. *Am J Sports Med*. 2007;35(4):659–668.
19. Vescovi JD. The menstrual cycle and anterior cruciate ligament injury risk: implications of menstrual cycle variability. *Sports Med*. 2011;41(2):91–101.
20. Wideman L, Montgomery MM, Levine BJ, Beynnon BD, Shultz SJ. Accuracy of calendar-based methods for assigning menstrual cycle phase in women. *Sports Health*. 2013;5(2):143–149.
21. Lenton EA, Landgren BM, Sexton L, Harper R. Normal variation in the length of the follicular phase of the menstrual cycle: effect of chronological age. *Br J Obstet Gynaecol*. 1984;91(7):681–684.
22. Chatterton RT, Mateo ET, Hou N, et al. Characteristics of salivary profiles of oestradiol and progesterone in premenopausal women. *J Endocrinol*. 2005;186(1):77–84.
23. DeSouza MJ, Toombs RJ, Scheid JL, O'Donnell E, West SL, Williams NI. High prevalence of subtle and severe menstrual disturbances in exercising women: confirmation using daily hormone measures. *Hum Reprod*. 2010;25(2):491–503.
24. Scarvell JM, Smith PN, Refshauge KM, Galloway HR, Woods KR. Association between abnormal kinematics and degenerative change in knees of people with chronic anterior cruciate ligament deficiency: a magnetic resonance imaging study. *Aust J Physiother*. 2005;51(4):233–240.
25. Sallis JF, Haskell WL, Wood PD, et al. Physical activity assessment methodology in the Five-City Project. *Am J Epidemiol*. 1985;121(1):91–106.
26. Gross LD, Sallis JF, Buono MJ, Roby JJ, Nelson JA. Reliability of interviewers using the Seven-Day Physical Activity Recall. *Res Q Exerc Sport*. 1990;61(4):321–325.
27. Klein LC, Bennett JM, Whetzel CA, Granger DA, Ritter FE. Caffeine and stress alter salivary alpha-amylase activity in young men. *Hum Psychopharmacol*. 2010;25(5):359–367.

28. Schwartz EB, Granger DA, Susman EJ, Gunnar MR, Laird B. Assessing salivary cortisol in studies of child development. *Child Dev*. 1998;69(6):1503–1513.
29. Bean JA, Leeper JD, Wallace RB, Sherman BM, Jagger H. Variations in the reporting of menstrual histories. *Am J Epidemiol*. 1979;109(2):181–185.
30. Small CM, Manatunga AK, Marcus M. Validity of self-reported menstrual cycle length. *Ann Epidemiol*. 2007;17(3):163–170.
31. Cole LA, Ladner DG, Byrn FW. The normal variabilities of the menstrual cycle. *Fertil Steril*. 2009;91(2):522–527.
32. Creinin MD, Keverline S, Meyn LA. How regular is regular? An analysis of menstrual cycle regularity. *Contraception*. 2004;70(4): 289–292.
33. Landgren BM, Unden AL, Diczfalusy E. Hormonal profile of the cycle in 68 normally menstruating women. *Acta Endocrinol (Copenh)*. 1980;94(1):89–98.
34. Shultz SJ, Wideman L, Montgomery MM, Levine BJ. Some sex hormone profiles are consistent over time in normal menstruating females: implications for sports injury epidemiology. *Br J Sports Med*. 2011;45(9):735–742.
35. DeSouza MJ, Miller BE, Loucks AB, et al. High frequency of luteal phase deficiency and anovulation in recreational women runners: blunted elevation in follicle-stimulating hormone observed during luteal-follicular transition. *J Clin Endocrinol Metab*. 1998;83(12): 4220–4232.

*Address correspondence to Bruce D. Beynnon, PhD, Department of Orthopaedics and Rehabilitation, University of Vermont, Burlington, Stafford Hall, 4th Floor, 95 Carrigan Drive, Burlington, VT 05405. Address e-mail to bruce.beynnon@uvm.edu.*