Center for Studies in Demography and Ecology

Ovulation Detection Methods for Urinary Hormones: Precision, Daily and Intermittent Sampling, and a Combined Hierarchical Method

by

Kathleen A. O'Connor
CSDE
& Department of Anthropology
University of Washington
Seattle, WA 98195

Eleanor Brindle
CSDE Biodemography Core
University of Washington
Seattle, WA 98195

Rebecca C. Miller
Drexel University College of Medicine
Philadelphia, PA 19129

R.J. Ferrell
Center for Population and Health
Georgetown University
Washington DC 20057

Jane B. Shofer
CSDE Statistical Core
University of Washington
Seattle, WA 98195

Nancy A. Klein
Seattle Reproductive Medicine
Seattle, WA 98109

Michael R. Soules
Seattle Reproductive Medicine
Seattle, WA 98109

Darryl J. Holman
CSDE
& Department of Anthropology
University of Washington
Seattle, WA 98195

P.K. Mansfield
Department of Women's Studies
& TREMIN Research Project on Women's Health
Pennsylvania State University,
University Park PA 16802

J.W. Wood
Department of Anthropology,
& Population Research Institute
Pennsylvania State University
University Park, PA 16802

Title: ***Ovulation Detection Methods for Urinary Hormones: Precision, Daily and Intermittent Sampling, and a Combined Hierarchical Method***

Running Title:

Ovulation Detection Using Urinary Hormones

Authors and Institutions:

K.A. O'Connor[1,2*]
E. Brindle[1*]
R.C. Miller[3]
J.B. Shofer[1]
R.J. Ferrell[4]
N.A. Klein[5]
M.R. Soules[5]
D.J. Holman[1,2]
P.K. Mansfield [6]
J.W. Wood[7]

[1] Center for Studies in Demography and Ecology, University of Washington, Seattle WA 98195
[2] Department of Anthropology, University of Washington, Seattle WA 98195
[3] Drexel University College of Medicine, Philadelphia, PA 19129
[4] Center for Population and Health, Georgetown University, Washington DC 20057
[5] Seattle Reproductive Medicine, Seattle, WA 98109
[6] Department of Women's Studies and TREMIN Research Project on Women's Health, Pennsylvania State University, University Park PA 16802
[7] Department of Anthropology, and Population Research Institute, Pennsylvania State University, University Park, PA 16802

*** These authors contributed equally to this manuscript**

Total Number of Pages: 35
Number of Figures: 1
Number of Tables: 6

Keywords: PDG/E1G/E1C/FSH/LH

Corresponding Author:
Kathleen A. O'Connor
Department of Anthropology
Box 353100
University of Washington
Seattle, Washington 98195
phone: (206) 543-9605; fax: (206) 543-3285
email: oconnork@u.washington.edu

Date: 11/18/2005

ABSTRACT

BACKGROUND: We evaluate the performance of ovulation detection methods and present new approaches, including evaluation of methods for precision, combining multiple markers into a hierarchical system, and using ovulation markers in intermittent sampling designs.

METHODS: With serum LH surge day as the "gold standard" of ovulation, we estimated accuracy and precision of ovulation day algorithms using 30 ovulatory menstrual cycles with daily urinary and serum hormones and transvaginal ultrasound. Sensitivity and specificity for estimating the presence of ovulation were tested using visually-assessed ovulatory (30) and anovulatory (22) cycles.

RESULTS: Sensitivity and specificity ranged between 70% and 100% for estimating presence of ovulation with twice-per-cycle, weekly, twice weekly, every-other-day and daily specimens. A combined hierarchical method estimated ovulation day using daily specimens within ± 2 days of the gold standard in 93% of cases. Accuracy of estimating ovulation day within ± 2 days using intermittent sampling ranged from 40% (weekly sampling) to 77% (every-other-day).

CONCLUSIONS: A combined hierarchical algorithm using precise and accurate markers allows maximal use of available data for efficient and objective identification of ovulation using daily specimens. In intermittent sampling designs, the presence but not the timing of ovulation can be estimated with good sensitivity, specificity, and accuracy.

INTRODUCTION

A variety of urinary hormone-based methods are available to detect the presence or absence and timing of ovulation, and they are increasingly used in large-scale and long-term studies of reproduction (e.g.(Burger *et al.,* 2005) (Liu *et al.,* 2004) (Wilcox *et al.,* 2004) (Miro *et al.,* 2004) (Santoro *et al.,* 2004) (Joseph-Horne *et al.*, 2002)). These methods thus merit both ongoing development and critical evaluation. In this paper we evaluate the performance of existing methods and present new approaches, including evaluation of ovulation marker methods for precision, combining multiple ovulation markers into a hierarchical system, and the use of ovulation markers in intermittent sampling designs.

Published methods for estimating the day of ovulation are usually evaluated for accuracy (Li *et al.,* 2002) (Santoro *et al.,* 2003) (Baird *et al.,* 1991) but none have been systematically examined for precision. Accuracy is a measure of how close a marker is to the day of ovulation. A marker may occur, on average, close to ovulation, and therefore be accurate, but we also need to know how precise it is; that is, whether there is high variation among women or cycles in the relationship of the marker with the day of ovulation. Ideally, we would like markers that are both accurate and precise. In the absence of a marker that is both accurate and precise, we should favor precise markers over accurate ones, because bias resulting from inaccuracy can be corrected, whereas bias from imprecision cannot. To facilitate choosing precise markers for estimating day of ovulation, we evaluated a range of urinary and serum hormone markers of ovulation for precision.

Technology in recent decades has facilitated the assay of specimens for multiple hormones. Research examining multiple hormones from daily specimens not only gives us the opportunity for a more comprehensive picture of reproductive function, but also provides us with

multiple markers of ovulation. However, few researchers have taken advantage of this in a systematic fashion. Dunson and colleagues used two urinary-based methods (Baird's day of luteal transition (Baird *et al.,* 1991) and the day of the luteinizing hormone surge) to estimate day of ovulation, and they point out that this approach improves accuracy and enables one to maintain sample size even when one of the markers, such as the easily missed LH surge, is not available in individual cases (Dunson *et al.,* 2001). We extend this approach to develop a hierarchical combination algorithm that takes advantage of the best indicators preferentially when applied to individual cases, and only uses less precise measures when better markers are unavailable.

Research sampling designs should be geared toward seeking a balance between minimizing subject collection burden and project costs, and providing sufficient data for reliable estimation of ovulation status and timing. Although several methods for estimating the presence or absence of ovulation have been evaluated for sensitivity and specificity using daily hormone data (Santoro *et al.,* 2003) (Baird *et al.,* 1991) (McConnell *et al.,* 2002) (Kassam *et al.,* 1996), they have not been evaluated for use with intermittent sampling designs. Similarly, methods for estimating the day of ovulation have been examined for accuracy using daily data (e.g. (Li *et al.,* 2002) (Santoro *et al.,* 2003) (Kassam *et al.,* 1996) (Baird *et al.,* 1995) but not intermittent data. To the best of our knowledge, there are no published, validated algorithms for estimating the presence or absence or day of ovulation using intermittent sampling designs. We therefore examined the performance of a range of urinary hormone-based methods for use with intermittent data collection strategies, including twice per cycle, once per week, twice per week and every-other-day.

Our goals were thus to 1) identify ovulation markers with the highest sensitivity, specificity, accuracy and precision; 2) combine the best methods into a hierarchical algorithm to maximize use of available data in studies using multiple hormone indicators; and 3) evaluate the performance of ovulation detection markers for intermittent sampling designs. In addition to providing methods for efficient and objective evaluation of ovulation, this information will facilitate the use of the best available methods for different research objectives, and can also enable researchers to account for error when estimating ovulatory status (e.g. (Dunson *et al.,* 2001) (Liu *et al.,* 2004)).

We used two data sets to achieve our goals. The first includes daily urinary PDG, E1G, LH and FSH data for 58 visually-assessed ovulatory and anovulatory menstrual cycles. The second includes daily serum and urine measurements of four reproductive hormones across the menstrual cycle and daily mid-cycle transvaginal ultrasound (TVU) measures from 30 ovulatory cycles.

MATERIALS AND METHODS

Participants and Specimens

Hormone results and menstrual diary data from two studies were used to develop and test urinary markers of ovulation. The first data set was used to evaluate methods for presence or absence of ovulation, and came from the Biodemographic Models of Reproductive Aging (BIMORA) project, a five-year, prospective study of the hormonal and menstrual cycle characteristics of reproductive aging (Ferrell *et al.,* 2005). Participants provided written informed consent and received compensation. The institutional review boards of Georgetown University, the University of Utah, the Pennsylvania State University, and the University of Washington approved all procedures. From January 15-July 14 in each of the five study years,

156 women collected daily first morning urine specimens and provided information on menstrual bleeding, prescription and over-the-counter supplement and drug use, and health conditions and medical procedures. We excluded data collected during and three months following any event that was known to affect menstruation or hormone levels (Ferrell *et al.,* 2005). A small subset of BIMORA data, 1,740 specimens from 58 menstrual cycles, is used in the present study. For this subset, we used only complete menstrual cycles that fell within one of the five six-month urine specimen collection windows.

In the second data set, 799 daily urine and serum specimens were collected across one menstrual cycle from 30 US women (13 aged 20-25 years and 17 aged 40-45) for a study on reproductive aging. Monetary compensation was provided, participants provided written informed consent, and all procedures were approved by the Institutional Review Board of the University of Washington. Participants were normally cycling, in good health, and not using medications or hormones. Daily morning blood specimens were obtained by venipuncture, beginning with the first day of menstrual bleeding and continuing until the first day of menstrual bleeding of the subsequent cycle. Daily TVU monitoring began in the mid-follicular phase and continued until follicle collapse was observed. Serum specimens were immediately assayed, all cycles were confirmed ovulatory by TVU, and day of follicle collapse was determined in 27 of the 30 cycles. Urine specimens were taken daily in the clinic, usually at the same time as blood collection, and thawed two years later for assay.

Assays

All serum specimens were assayed for estradiol (E2), luteinizing hormone (LH), and follicle stimulating hormone (FSH), but serum progesterone (P4) was measured only in the luteal phase. The RIA for E2 (ICN Biomedicals, Irvine, CA, USA) had inter- and intra-assay CVs of

16% and 7%. The RIA for P4 (Diagnostic Systems Laboratories, Webster, TX, USA) had inter- and intra-assay CVs of 13% and 11%. Solid-phase two-site immunofluorometric assays (IFMA) (Delfia; Wallac, Turku, Finland) were used to estimate the serum levels of intact LH and FSH. The LH IFMA cross-reacts less than 1% with FSH, and the inter- and intra-assay CVs were 2.8% and 4.7% respectively. The FSH IFMA cross reacts less than 1% with LH, and the inter- and intra-assay CVs were 2.3% and 4.6% respectively.

Competitive enzyme immunoassays (EIAs) were used to assay urine specimens for pregnanediol glucuronide (PDG), estrone glucuronide (E1G), and estrone conjugates (E1C). The PDG, E1C and E1G EIAs are described in detail elsewhere (O'Connor *et al.,* 2003) (O'Connor *et al.,* 2004). The inter- and intra-assay coefficients of variation were 10.3% and 9.2% respectively for the PDG EIA (O'Connor *et al.,* 2003), 3.2% and 3.6% for the E1G EIA (O'Connor *et al.,* 2004), and 10.9% and 7.3% for the E1C EIA (O'Connor *et al.,* 2003). Urinary and serum concentrations were highly correlated: r = 0.98 for PDG-progesterone (O'Connor *et al.,* 2003), r = 0.94 for E1G-estradiol (O'Connor *et al.,* 2004), and r = 0.93 for E1C-estradiol (O'Connor *et al.,* 2003).

Sandwich immunoenzymometric assays (IEMAs) were used to measure beta-LH and beta-FSH in urine; these assays are described in detail elsewhere (Brindle *et al.,* submitted). The inter- and intra-assay coefficients of variation for the LH assay were 6.8% and 3.7% respectively. The inter- and intra-assay coefficients of variation for the FSH assay were 3.7% and 4.1% respectively. Urinary LH and FSH values were highly correlated with serum hormone results: pearson correlations between urine and serum values for 30 averaged cycles were 0.86 for FSH and 0.95 for LH (Brindle *et al*., submitted).

All specimens, calibrators and controls were run in duplicate, and hormone concentrations were estimated from optical density using a four parameter logistic model in Biolinx 2.0 Software (Dynatech Laboratories, Inc., Chantilly, VA, USA). All urinary hormone values were adjusted by specimen specific gravity, using a population mean specific gravity of 1.020 (Miller *et al.,* 2004).

Estimating Presence or Absence of Ovulation

The ability of algorithms to detect the presence or absence of ovulation was evaluated using 1) sensitivity (the ability to detect a true positive [ovulation], defined as the number of cycles classified as ovulatory/true number of ovulatory cycles); 2) specificity (the ability to detect a true negative [anovulation], defined as the number of cycles classified anovulatory/true number of anovulatory cycles), and 3) the percent of cycles correctly classified by a method (defined as the number of cycles correctly classified/total number of cycles). The Kassam (Kassam *et al.,* 1996) PDG moving averages interval method, the Baird (Baird *et al.,* 1991) day of luteal transition (DLT) method, and mid-cycle LH and FSH peaks (Aedo *et al.,* 1976) (Kesner *et al.,* 1998) (Santoro *et al.,* 2003) were tested.

The Kassam (Kassam *et al.,* 1996) method uses a ratio of daily PDG to the minimum 5-day moving average PDG to identify a sustained PDG rise, defined as three consecutive days of a ratio of target day PDG to the 5 day running average ≥ three times the baseline in complete cycles (cycle method) or across time intervals (interval method) when cycle day is uncertain. The Baird et al. (Baird *et al.,* 1991) algorithm identifies the presence and timing of ovulation using rules for finding an abrupt change in the ratio of estrogen to progesterone. To identify urinary LH and FSH peak days, we modified existing methods which identify significant LH peaks (Aedo *et al.,* 1976) (Kesner *et al.,* 1998) (Santoro *et al.,* 2003). Rather than choosing the

day with the highest concentration as evidence of ovulation, we considered the timing of the LH and FSH peaks to distinguish those occurring around the time of menses (e.g. (Miro and Aspinall 2005)) from those at mid-cycle. Additionally, we specified that an LH or FSH peak be followed by a rise in PDG for a cycle to be deemed ovulatory. Peak days were identified as those with the highest LH or FSH value ≥ the average LH or FSH value for the entire menstrual segment plus two standard deviations of all values for the menstrual segment. Only peaks occurring ≤ 4 days before, on the day of, or one day after the first day of a sustained PDG rise (as defined by the Kassam algorithm) were considered.

The above four methods (Kassam, Baird, LH, FSH) were applied to 58 menstrual cycles from the BIMORA project, for which we had daily urines and menstrual diaries. The cycles were selected at random and independently evaluated for ovulation status by two researchers with extensive experience interpreting urinary reproductive hormone profiles. Cycles were assessed by visual inspection of the patterns of E1G, PDG, LH, and FSH. Both experts were in agreement on the classifications: 30 were ovulatory, 22 were anovulatory, and 6 were indeterminate. Neither researcher was aware of the age or reproductive history of the women contributing the segments. Algorithm performance was evaluated by estimating sensitivity and specificity of the four methods using the visual classification of presence or absence of ovulation as the "gold standard."

Estimating Day of Ovulation

The ability of ovulation algorithms to estimate the day of ovulation was examined for a range of markers using the data set of 30 ovarian cycles confirmed ovulatory by TVU. Performance of each marker was evaluated by precision (consistency across women or cycles in timing relative to ovulation) and accuracy (proximity to ovulation day). For each of the methods

discussed below, accuracy was assessed by how close the mean difference between a marker and the gold standard (marker minus gold difference or "MMGD") was to zero. Precision of each marker was assessed as 1) the standard deviation of the MMGD, and 2) the percentage of estimates on and within ±1 and ±2 days of the mode of the MMGD. Markers designated as precise have the smallest MMGD standard deviation and the highest percentage of estimates falling on or within ±1 or ±2 days of the MMGD mode. Cases for which a marker was not available were not included in estimates of either accuracy or precision.

We selected the day of the serum LH surge as our gold standard; the day of the serum LH surge is considered the day of ovulation, and occurs, on average, 1 day earlier than follicle collapse observed by ultrasound (WHO 1980) (Queenan *et al.,* 1980) (Pauerstein *et al.,* 1978).

Several widely used methods for identifying ovulation day were evaluated: two versions of Baird's day of luteal transition (DLT) method, one using E1G and one using E1C (Baird *et al.,* 1991); the Waller (Waller *et al.,* 1998) method; the mid-cycle peak day of urinary LH (Kesner *et al.,* 1998), and the mid-cycle peak day of urinary FSH (Li *et al.,* 2002). Baird and colleagues' (Baird *et al.,* 1991) DLT algorithm identifies the day of ovulation as the 2nd day in a 5 day sequence where there is a 40% or greater decline in the ratio of estrogen to progesterone between the first and the last two days of the sequence. The Waller et al. (Waller *et al.,* 1998) method is a modification of the Baird DLT method, with more specific criteria for identifying ovulation day based on the peak and subsequent change in the estrogen to progesterone ratio, and a modification of the PDG ratio to accommodate very high and low levels of PDG. We modified existing methods for identifying ovulatory LH and FSH peaks to consider the magnitude and timing of the peak to distinguish mid-cycle peaks from those occurring around menses (Miro and Aspinall 2005). The peak day had the highest LH or FSH value ≥ the average LH or FSH value

for the entire menstrual cycle plus two standard deviations of all values for the cycle. Only peaks occurring within 4 days before, on the day of, or one day after the first day the Kassam PDG ratio was ≥ 3 for at least 3 consecutive days were considered.

We also evaluated methods less commonly used—the peak day of urinary E1G and the peak day of urinary E1C (Li *et al.,* 2002) (Baird *et al.,* 1991)—or not previously examined—Kassam's (Kassam *et al.,* 1996) urinary PDG-based interval and cycles method—for their ability to identify ovulation day. The Kassam method for identifying a sustained rise in PDG was modified to estimate the day of ovulation: we used the first day of the sustained rise in PDG as the day of ovulation. The urinary E1G and E1C peaks were defined as the highest values in a menstrual segment.

We also examined the relationships of serum FSH, LH and E2 peaks with the gold standard. The same peak identification methods used for urinary peaks of these hormones were used with the serum data.

Combination Hierarchical Algorithm

The best-performing urinary hormone markers were combined to create a hierarchical method for determining presence or absence and day of ovulation. The combined method was constructed using four criteria: 1) the presence or absence of ovulation was estimated using the method with the highest sensitivity and specificity; 2) if ovulation occurred, the marker with the highest precision and accuracy was used to estimate day of ovulation; 3) the hierarchy favored the most precise and accurate method for determining ovulation day, but would compensate for its absence on a case-by-case basis by using the second most accurate or precise method in the hierarchy; if an individual case did not have data for the second-choice method, then the third most accurate or precise method in the hierarchy was used, and so on; 4) the individual methods

in the hierarchy had to be amenable to automation. We considered precision more important than accuracy, so markers consistently the same number of days from the gold standard were ranked higher than those occurring nearer the gold standard but which had high variability from cycle to cycle.

Estimating Presence or Absence of Ovulation with Intermittent Specimen Sampling Designs

Intermittent sampling designs using every-other-day, twice-weekly, weekly, and twice-per-menstrual cycle collection protocols were simulated in our data sets. Every-other-day collection was simulated by randomly selecting either the first or second day of data available for each cycle, and then every second day through the rest of the cycle. For twice-weekly sampling, two situations were simulated. In the first, the first day of specimen collection was the first day of menstrual bleeding, and the next day of collection was three days later, followed by a collection four days later, followed by a collection three days later, and so on. In the second simulated situation, specimens were collected every Monday and Thursday, beginning on the first Monday in a menstrual cycle. To simulate weekly collection, similar scenarios were used: beginning collection on the first day of menses, or on a Monday. To simulate twice-per-menstrual cycle sampling, we used the 7^{th} and 21^{st} day of each cycle, assuming an average 28-day cycle length, given that specimen collection days would need to be decided without knowing completed cycle length in advance.

These sampling designs were simulated using both data sets combined, for a total of 88 menstrual cycles. Ovulation algorithms were applied to determine the presence or absence of ovulation in each menstrual cycle, and performance was assessed by estimating sensitivity and specificity against the "gold standard" determination made using the most sensitive and specific algorithm with daily data.

The ovulation algorithms were designed for use with daily data, and had to be modified for intermittent sampling. We evaluated only the Kassam and Baird methods; these are both based on a rise in PDG, which is the only indicator that spans enough days to accommodate intermittent sampling designs. Our modifications were designed to balance reduced observations with stringent criteria indicative of ovulation or its absence. The modifications are shown in **Table I**, along with the reasoning for each modification.

TABLE I

Estimating Day of Ovulation with Intermittent Specimen Sampling Designs

To estimate the day of ovulation with data collected weekly, twice-weekly, and every-other-day, variations of the basic principles of the Kassam algorithm were evaluated for precision and accuracy. For example, for weekly data we examined the first day the Kassam ratio exceeded 3 as well as up to 4 days before and up to 5 days after in order to identify which algorithm had the highest precision and accuracy. The denominator of the ratio also varied by sampling design: it was simply the minimum PDG value in a menstrual segment for weekly and twice-weekly sampling, but was calculated using a running 3 sample average for every-other-day data.

Accuracy of the different variations of the method was evaluated using the mean MMGD and the percent of cases in which the day of ovulation was estimated within ± 2 and ± 4 days of the serum LH peak. Precision was evaluated using the MMGD SD. The thirty cycles with serum LH and TVU were used for these tests.

RESULTS

The mean and standard deviation of cycle lengths for each data set are shown in **Table II**. The anovulatory and indeterminate cycles had wider variation in cycle length than the ovulatory cycles.

TABLE II

Estimating Presence or Absence of Ovulation With Daily Data

Sensitivity, specificity and misclassification rate for each of the algorithms tested for determining the presence or absence of ovulation using daily specimens are shown in **Table III**. TABLE III Of four algorithms applied to the 52 classifiable menstrual segments from the BIMORA study, the Kassam algorithm using the intervals method (Kassam *et al.,* 1996) performed best overall in terms of sensitivity (100%), specificity (100%), and misclassification (0%) (**Table III**). TABLE III Of the 6 cycles visually classified as indeterminate, 5 were classified anovulatory and one ovulatory by the Kassam algorithm. Examples of the steroid and gonadotropin hormone profiles of ovulatory, anovulatory and indeterminate cycles are shown in **Figure 1**. Visually indeterminate cycles lack a clear and sustained rise in PDG, and tended to be classified as anovulatory by each of the algorithms. Figure 1 The visually indeterminate cycle in **Figure 1C** was assigned anovulatory status by the Kassam method. Figure 1

Estimating Day of Ovulation with Daily Data

Table IV presents measures of precision for each of the different markers: the standard deviation of the mean MMGD, the MMGD mode, and the percent of cases where a marker falls on or within 1 and 2 days of the MMGD mode. TABLE IV Only cycles where the indicator being tested was present were included in the denominator for calculating percentage of cases falling on or within 1 and 2 days of the mode; absence of the indicator would otherwise have been treated as a method failure, unduly lowering the precision estimate. Accuracy is also presented in **Table IV**, TABLE IV as the mean MMGD, representing the proximity of a marker to the gold standard.

TABLE IV

Table IV shows that the means of the E1C and E1G peaks are close in time to the day of serum LH peak (MMGD=0.87 days for E1C and 0.30 days for E1G), and are therefore accurate, but the standard deviations are large (MMGD SD = 5.99 days for E1C and 2.84 days for E1G)

indicating high inter-subject variability, and thus poor precision. Given their low precision, E1C and E1G peaks were not considered in subsequent analyses, or considered for the hierarchical method. The day of PDG rise identified using the Kassam method is not particularly accurate—the first day of the PDG rise occurs on average three days following the serum LH peak (MMGD = 3.48 days for the Kassam cycle method, MMGD = 3.17 days for the Kassam interval method)—but it has reasonable precision (MMGD SD = 1.92 for Kassam cycle method; MMGD SD= 1.60 for Kassam interval method). The timing of the rise was considered consistent enough from woman to woman to allow us to take the three day lag into account in subsequent analyses using the Kassam method to estimate day of ovulation. Urinary LH, FSH, Baird DLT and the Waller methods were all reasonably accurate (mean MMGDs close to zero), and precise (low SDs). Serum E2 is actually fairly precise (77% of cases fell on the mode); the relatively large MMGD SD (1.82 days) for E2 primarily reflects one case with an 8 day deviation from the LH surge.

The mean of the mean deviations between day of serum LH surge and all other markers in **Table IV** was 0.71 days (SD=0.99). The mean of the mean deviations between day of follicle collapse and all other markers in **Table IV** was 0.29 days (SD=0.97). Follicle collapse is closer in time to the marker events, but not significantly so, indicating that either serum LH surge or follicle collapse can be used to represent day of ovulation.

TABLE IV

TABLE IV

The results presented in **Tables III** and **IV** guided the construction of a combined hierarchical method. In this method, the first step was to evaluate whether or not a cycle was ovulatory. **Table III** supported the use of the Kassam (Kassam *et al.,* 1996) intervals method algorithm (100% sensitivity and specificity). If the Kassam algorithm identified a cycle as ovulatory, the next step was to identify the day of ovulation. The urinary LH peak, FSH peak

TABLES III, IV

TABLE III

and Baird DLT methods all had similar precision and accuracy. Using precision as a guide, we placed the LH peak first (MMGD SD=1.14) and the FSH peak second (MMGD SD=1.30) in the hierarchical algorithm. LH and FSH peaks were put ahead of the Baird DLT (MMGD SD=0.85 for E1G, MMGD SD=1.14 for E1C) because of the sometimes unusual ovulation days that can be assigned by the Baird DLT algorithm in the hormonally-irregular cycles that can occur as a result of reproductive aging (data not shown). The Baird algorithm worked equally well with both the E1C and E1G measures, and either of these can be used in the combined hierarchical method. The Kassam intervals method, with correction for the three day lag, had lower precision (MMGD SD=1.60) than LH, FSH and the Baird DLT and thus it was placed last in the hierarchy. Ovulation day using the Kassam method is identified as three days before the first day of three consecutive days that the ratio exceeded the threshold value of 3.

The combined hierarchical method had, overall, high accuracy (mean MMGD = 0.60) and precision (MMGD SD=1.13; 93% of estimates fell within 2 days of the MMGD mode) for predicting ovulation day (**Table IV**). We did not include the Waller method in the hierarchical method as it is a more complicated version of the Baird DLT method and had lower accuracy and precision than the Baird method.

TABLE IV

Intermittent Sampling

Table V shows results for six different intermittent sampling scenarios for estimating the presence or absence of ovulation. The modified Baird and Kassam algorithms performed similarly with respect to sensitivity for every-other-day (range = 98-100%), twice-weekly (range = 92-100%) and weekly (range = 78-93%) sampling, regardless of when collection began. Misclassification rates were below 10% for methods that had both sensitivities and specificities greater than 90%. Most of the algorithms were too sensitive to small PDG changes and therefore

TABLE V

had low specificity with every-other-day sampling (5-77%). However, requiring an elevation in PDG of three times the baseline for two consecutive samples improved specificity to 100% for every-other-day samples (using the Kassam 6 modification, see **Tables I and V**). Specificity also improved for weekly and twice-weekly sampling with different modifications to the Kassam algorithm, such as raising the threshold ratio from 3 to 4 (using the Kassam 4 and 5 modifications, see **Tables I and V)**. Twice per cycle collection had relatively low sensitivity and specificity (68-82%) for estimating the presence or absence of ovulation. The Kassam method variations generally had higher sensitivity and specificity and lower misclassification rates than the Baird method variations. Methods for determining presence or absence of ovulation had to be tailored to accommodate the collection scheme: no single method was identified that could be used to give the best results for all sampling scenarios.

TABLES I, V

TABLES I, V

Accuracy and precision for identifying ovulation day with intermittent data is shown in **Table VI**; only the best performing (highest accuracy and precision) algorithms are listed. Precision was roughly similar (MMGD SD ranging from 2.0 to 2.8 days) for weekly, twice-weekly and every-other-day sampling designs (**Table VI**). Accuracy of PDG-based estimates of the day of ovulation increased with increased sampling frequency: the MMGD declined from 2.1 to -0.4 as sampling frequency increased (**Table VI**). Twice-weekly and every-other-day methods estimated ovulation day within ±4 days in up to 90% of cases but only 63-77% were estimated within ±2 days. Weekly sampling estimated ovulation day within ±2 days in only 40-70% of cases (**Table VI**). For both twice-weekly and weekly sampling, beginning specimen collection on the first Monday of a menstrual cycle resulted in higher accuracy and precision than beginning collection on the first day of a menstrual cycle (the first day of menses).

TABLE VI

TABLE VI

TABLE VI

TABLE VI

DISCUSSION

The results of this work provide researchers with information for choosing urinary markers of ovulation that 1) are precise; 2) are well-characterized statistically; 3) make maximum use of available ovulation marker data, even in cases where different types of data are missing; 4) can be applied objectively; and 5) allow for choice of a research design that balances subject collection burden, project goals, and cost with appropriate data for estimation of the presence or absence and day of ovulation.

The PDG-based Kassam method, and modifications of it, had high sensitivity and specificity for estimating the presence or absence of ovulation in daily and intermittent sampling scenarios. The Kassam method had 100% specificity and sensitivity for determining whether ovulation occurred in a cycle with daily sampling. This method easily accommodated variability between subjects in magnitude of the PDG rise, and could also be applied with minimal modification to a range of sampling schemes. With intermittent sampling schemes, including every-other-day, twice weekly and weekly, the Kassam method modifications had sensitivities and specificities greater than 90% and misclassification rates of 10% or less. Even twice per cycle sampling had sensitivity and specificities from 68-82% with the Kassam method. While the Kassam method proved to be an effective basis, modifications specific to each collection scenario had to be made to optimize performance. This was a result of the need to balance data richness with the strictness of the rules in order to achieve high sensitivity and specificity. For less frequent sampling, more permissive rules were needed to optimize sensitivity. These rules were too relaxed, however, and compromised specificity in designs with higher sampling frequency. The Baird DLT method had high sensitivity but low specificity for most sampling designs; the low specificity was probably a result of the method's vulnerability to missing data (Baird *et al.,* 1991).

Other methods for estimating the presence or absence of ovulation tend not to be as sensitive or specific as the Kassam method with daily sampling (Santoro *et al.,* 2003). In our data, LH and FSH peak days were specific, but not sensitive, while Baird's DLT method was sensitive but not very specific. The Kassam method was chosen for estimating the presence or absence of ovulation in the combined hierarchical method because of the method's high sensitivity and specificity as well as robustness to missing data.

Although the Kassam method performed quite well with our daily data for estimating presence or absence of ovulation, slightly lower sensitivity and specificity were reported in its original description (Kassam *et al.,* 1996) and elsewhere (Santoro *et al.,* 2003;McConnell *et al.,* 2002). This might be attributable, in part, to our gold standard choice. Whereas Kassam and colleagues used <2 ng/mL serum progesterone as a gold standard of anovulation, taken once per week, and McConnell and colleagues used the urine LH peak as a gold standard of the presence of ovulation, we tested our algorithms for the presence or absence of ovulation against a gold standard of visually assessing cycles using measures of four hormones from daily urine specimens. Although our visual assessment included E1G, PDG, LH and FSH, the most telling visual indicator of ovulation was a sustained mid-cycle rise in PDG. Thus, given that the test and gold standard were based on similar (but not identical) criteria, our estimates of sensitivity and specificity may be overly optimistic. However, our objective was to effectively capture and standardize the PDG rise that is evident visually, even with intermittent sampling, using the Kassam algorithm.

We evaluated the performance of algorithms for the presence or absence of ovulation in a data set that included 25 cycles from perimenopausal women (out of a total of 58 cycles from the BIMORA project). Perimenopausal women may have hormonally disordered cycles with or

without ovulation, including elevated early follicular phase FSH, lack of LH or FSH peaks, and hyper- or hypo-estrogenism (Prior 1998). Many of the cycles in our sample showed one or more of these characteristics (for an example, **see Figure 1B**). Despite this, the Kassam PDG-based algorithm had nearly perfect sensitivity and specificity for discriminating ovulatory and non-ovulatory cycles. Our results are in agreement with Santoro and colleagues (Santoro *et al.,* 2003) in finding the Kassam method useful for evaluating not only normal cycles, but also the hormonally irregular cycles encountered in reproductive aging. In previous work we found that the Kassam algorithm also performed well in the disrupted cycles sometimes found in exercising women (McConnell *et al.,* 2002). We note, however, that the algorithms examined here were designed for use in women still experiencing menstrual cycles. It is possible that they may not perform similarly in women who are not cycling at all, such as post-menopausal women, in whom low-level fluctuations in steroid hormone levels may cause the algorithms to falsely indicate that ovulation has occurred. The chances of identifying a PDG rise three times the baseline that is associated only with random, low-level fluctuations will increase as interval length increases, as is expected in peri- or post-menopausal women. In these cases, it may be necessary to identify a minimum PDG level below which peaks identified by the Kassam method are clearly not biologically meaningful.

Figure 1

All the indictors we examined for estimating of day of ovulation were accurate within one day of ovulation, except for the Kassam algorithm which gave an ovulation day three days removed, on average, from the gold standard. We were able to correct for this bias by designating the day of ovulation as 3 days prior to the first of three consecutive days when the ratio was ≥ 3. Precision varied across urinary indicators, ranging from 0.85 to 5.99 days. The E1G and E1C peaks, despite being accurate to within one day of ovulation, were the least precise

indicators and were thus not considered for use in the combined hierarchical method. The remaining urinary indicators had acceptable precision, ranging from 0.85 to 1.92 days.

In the construction of the combined hierarchical method for estimating day of ovulation we valued precision over accuracy. If a cycle was found ovulatory by the Kassam method, the mid–cycle urine LH peak, if available, was used preferentially to estimate ovulation day. If the urinary LH peak was not available, then the mid-cycle urine FSH peak was used as the day of ovulation. If no LH or FSH peaks were available, then ovulation day was estimated using the Baird DLT method. If the Baird algorithm identified the cycle as anovulatory, or the ovulation day indeterminate, the last step of the hierarchy was to define the day of ovulation as three days before the first of three consecutive days on which the Kassam PDG:PDG baseline ratio was ≥ 3. Urinary LH and FSH peaks and the Baird DLT algorithm, though reasonably accurate and precise, are all vulnerable to small amounts of missing data (Baird *et al.,* 1995), in contrast to the Kassam method. Thus, the Kassam method is an important component of the combined hierarchical method. The combined hierarchical method had excellent accuracy—the MMGD mean was less than one day (0.60 days)— and excellent precision—93% of estimates of ovulation day in a sample of 30 cycles fell on or within two days of the MMGD mode, and the MMGD SD was 1.13 days. As an example of how the method accommodates missing data, when we applied the algorithm to 61 cycles classified as ovulatory (30 confirmed ovulatory by TVU, 30 visually-assessed ovulatory, 1 visually-assessed indeterminate) ovulation day was determined by the urine LH peak in 42 cycles, by the FSH peak in 4 cycles, by Baird DLT in 13 cycles, and by the Kassam method in 2 cycles.

Variants of the Kassam method were used to estimate the day of ovulation with intermittent data. The broad peak in PDG captured by this method lends itself to use in

intermittent sampling. Other methods were not evaluated given their vulnerability to missing data. The hierarchical combined method was also not useful for intermittent data because all of the indicators in that method, except for the PDG Kassam ovulation day component, were too vulnerable to missing data. Estimation of ovulation day with Kassam-based methods to within ± 4 days of ovulation had 90% accuracy for twice weekly and every-other-day sampling, but accuracy declined to around 70% for estimation to within ± 2 days of ovulation day. Thus, every-other-day and weekly sampling could be used if pinpointing the day of ovulation to within ±4 days is acceptable for the study design.

Beginning specimen collection on the first Monday of a menstrual cycle resulted in higher accuracy and precision than beginning collection on the first day of a menstrual cycle (the first day of menses) for estimating ovulation day with both twice-weekly and weekly sampling. This is a result of the fact that ovulation is not distributed randomly across the menstrual cycle, and tends to cluster between cycle days 8-15 (Wilcox *et al.*, 2000). The day of ovulation (day of serum LH surge) in the 30 cycles for which we had TVU clustered around days 13-14 of the menstrual cycle: the mean ovulation day was cycle day 14.3, the mode was cycle day 13, the median was cycle day 14 and the range spanned cycle days 7 to 25. Because ovulation day clustered so tightly around cycle days 13-14, and these cycle days fell between collection days, beginning collection on the first day of menses missed ovulation day in a substantial proportion of our data set in the twice-weekly and weekly sampling designs. In contrast, beginning collection on the first Monday of a cycle increased the probability of collection days coinciding with cycle days 13 and 14. We conclude that intermittent collection schemes for weekly and twice-weekly collection should avoid beginning specimen collection on the first day of menses.

A limitation of the work here is that the methods were evaluated on relatively small data sets—30 ovulatory cycles with serum, urine and TVU data, and 58 visually-assessed anovulatory and ovulatory BIMORA cycles. Our estimates of sensitivity, specificity, precision and accuracy may not be representative of how the markers might perform in other data sets. In particular, our data might reflect a limited range of variance across women for the estimates of precision. It is not possible to compare our estimates of accuracy with other studies, as each study used different gold standards: we used the serum LH peak, Baird and colleagues used the urinary LH peak (Baird *et al.,* 1991), Li and colleagues used follicle collapse (Li *et al.,* 2002), and Santoro and colleagues used visual assessment of evidence of luteal activity (Santoro *et al.,* 2003) as gold standards for estimating accuracy. A second and related limitation is that our estimates of precision, accuracy, sensitivity and specificity for methods that we modified or created are likely to be biased by the fact that we are reporting statistical performance for indicators using the data set which was used to develop or modify the indicators. Thus, it is possible that the indicators discussed here may not perform as well in other data sets.

The methods outlined here were developed for objective and automated identification of the presence and timing of ovulation. The results suggest that a combined hierarchical method has several advantages for daily samples: it has excellent sensitivity and specificity for estimating presence of ovulation; it has high accuracy and precision for estimating ovulation day; and it is robust to missing data, thereby maximizing use of a data set. We found that the presence or absence of ovulation can be estimated with PDG-based methods with good sensitivity and specificity in intermittent sampling designs. In particular, every-other-day, twice-weekly and weekly specimens perform well. These results should be useful for reducing subject burden and project cost in studies in which the outcome measure is presence or absence of

ovulation. Intermittent sampling designs are less useful for estimating day of ovulation; for the best accuracy and precision, daily sampling is necessary.

ACKNOWLEDGEMENTS

We thank J. Aranda, C. Mar, D. Schechter and K. Wander for their contributions to this work, and we are especially grateful to the BIMORA participants. This work was funded by: NICHD 1RO1HD034159; NIA 1RO1AG015141; NIA RO1AG14579; NICHD R24HD042828; the Center for Studies in Demography and Ecology, University of Washington, and the Center for Population and Health, Georgetown University.

REFERENCES

Burger HG, Robertson DM, Baksheev L, Collins A, Csemiczky G and Landgren B (2005) The relationship between the endocrine characteristics and the regularity of the menstrual cycles in the approach to menopause. Menopause 12:267-274.

Liu Y, Gold EB, Lasley BL and Johnson WO (2004) Factors affecting menstrual cycle characteristics. Am J Epidemiol 160:131-140.

Wilcox AJ, Baird DD, Dunson DB, McConnaughey DR, Kesner JS and Weinberg CR (2004) On the frequency of intercourse around ovulation: Evidence for biological influences. Hum Reprod 19:1539-1543.

Miro F, Parker SW, Aspinall LJ, Coley J, Perry PW and Ellis JE (2004) Origins and consequences of the elongation of the human menstrual cycle during the menopausal transition: The freedom study. J Clin Endocrinol Metab 89:4910-4915.

Santoro N, Lasley BL, McConnell D, Allsworth J, Crawford SL, Gold EB, Finkelstein JS, Greendale GA, Kelsey J, Korenman SG *et al.,* (2004) Body size and ethnicity are associated with menstrual cycle alterations in women in the early menopausal transition:

The study of women's health across the nation (swan) daily hormone study. J Clin Endocrinol Metab 89:2622-2631.

Joseph-Horne R, Mason H, Batty S, White D, Hillier S, Urquhart M and Franks S (2002) Luteal phase progesterone excretion in ovulatory women with polycycstic ovaries. Hum Reprod 17:1459-1463.

Li H, Chen J, Overstreet JW, Nakajima ST and Lasley BL (2002) Urinary follicle-stimulating hormones peak as a biomarker for estimating the day of ovulation. Fertil Steril 77:961-966.

Santoro N, Crawford SL, Allsworth JE, Gold EB, Greendale G, Korenman SG, Lasley BL, McConnell D, McGaffigan P, Midgely R *et al.,* (2003) Assessing menstrual cycles with urinary hormone assays. Am J Physiol Endocrinol Metab 284:E521-E530.

Baird DD, Weinberg CR, Wilcox AJ, McConnaughey DR and Musey PI (1991) Using the ratio of urinary oestrogen and progesterone metabolites to estimate day of ovulation. Stat Med 10:255-266.

Baird DD, McConnaughey R, Weinberg CR, Musey PI, Collins DC, Kesner JS, Knecht EA and Wilcox AJ (1995) Application of a method for estimating day of ovulation using urinary estrogen and progesterone metabolites. Epidemiology 6:547-550.

Dunson D, Weinberg C, Baird D, Kesner J and Wilcox A (2001) Assessing human fertility using several markers of ovulation. Stat Med 20:965-978.

McConnell HJ, O'Connor KA, Brindle E and Williams NI (2002) Validity of methods for analyzing urinary steroid data to detect ovulation in athletes. Med Sci Sports Exerc 34:1836-1844.

Liu Y, Johnson W, Gold E and Lasley B (2004) Bayesian analysis of risk factors for anovulation. Stat Med 23:1901-1919.

Ferrell RJ, O'Connor KA, Rodríguez G, Gorrindo T, Holman DJ, Brindle E, Miller RC, Schechter D, Korshalla L, Simon JA*et al.,* (2005) Monitoring reproductive aging in a five year prospective study: Aggregate and individual changes in steroid hormones and menstrual cycle lengths with age. Menopause 12:567-577.

O'Connor KA, Brindle E, Holman DJ, Klein NA, Soules MR, Campbell KL, Kohen F, Munro CJ, Shofer JB, Lasley BL *et al.,* (2003) Urinary estrone conjugate and pregnanediol-3-glucuronide enzyme immunoassays for population research. Clin Chem 49:1139-1148.

O'Connor KA, Brindle E, Shofer JB, Miller RC, Klein NA, Soules MR, Campbell KL, Mar C and Handcock MS (2004) Statistical correction for non-parallelism in a urinary enzyme immunoassay. J Immunoassay Immunochem 25:259-278.

Miller RC, Brindle E, Holman DJ, Shofer JB, Klein NA, Soules MR and O'Connor KA (2004) Comparison of specific gravity and creatinine methods for normalizing urinary reproductive hormone concentrations. Clin Chem 50:924-932.

Aedo AR, Landgren BM, Cekan Z and Diczfalusy E (1976) Studies on the pattern of circulating steroids in the normal menstrual cycle. 2. Levels of 20alpha-dihydroprogesterone, 17-hydroxy-progesterone and 17-hydroxypregnenolone and the assessment of their value for ovulation prediction. Acta Endocrinol 82:600-616.

Kesner JS, Knecht EA, Krieg EF, Wilcox AJ and O'Connor J (1998) Detecting pre-ovulatory luteinizing hormone surges in urine. Hum Reprod 13:15-21.

Miro F and Aspinall LJ (2005) The onset of the initial rise in follicle-stimulating hormone during the human menstrual cycle. Hum Reprod 20:96-100.

WHO (1980) Temporal relationships between ovulation and defined changes in the concentration of plasma estradiol-17ß, luteinizing hormone, follicle-stimulating hormone, and progesterone. Am J Obstet Gynecol 138:383-390.

Queenan JT, O'Brien GD, Bains LM, Simpson J, Collins WP and Campbell S (1980) Ultrasound scanning of ovaries to detect ovulation in women. Fertil Steril 34:99-105.

Pauerstein CJ, Eddy CA, Croxatto HD, Hess R, Siler-Khodr and Croxatto HB (1978) Temporal relationships of estrogen, progesterone, and luteinizing hormone levels to ovulation in women and infrahuman primates. American Journal of Obstetrics and Gynecology 130:876-886.

Waller K, Swan SH, Windham GC, Fenster L, Elkin EP and Lasley B (1998) Use of urine biomarkers to evaluate menstrual function in healthy premenopausal women. Am J Epidemiol 147:1071-1080.

Prior JC (1998) Perimenopause: The complex endocrinology of the menopausal transition. Endocr Rev 19:397-428.

Wilcox AJ, Dunson D and Baird DB (2000) The timing of the "fertile window" in the menstrual cycle: Day specific estimates from a prospective study. BMJ 321:1259-1262.

Table I. Modifications to the Kassam and Baird methods for estimating the presence or absence of ovulation in intermittent sampling designs

Method modified	Sampling design	Modification	Justification
Kassam (1)	Twice per cycle Weekly Twice Weekly	Baseline (ratio denominator) is minimum value for the cycle; ratio has to be >3 on day following minimum value	Not enough data for a 5-day baseline
Kassam (2)	Every-other-day Twice Weekly	Baseline (ratio denominator) is minimum value for the cycle; ratio has to be >3 for two consecutive samples, disregard any ratio >3 on 1st day of collection.	Not enough data for a 5-day baseline; fewer days exceeding threshold to compensate for fewer observations
Kassam (3)	Every-other-day Twice Weekly	Baseline calculated using 3, rather than 5 specimen running average, and threshold of ratio >3 exceeded for 1, not 3, days, disregard any ratio >3 on 1st day of collection	Not enough data for a 5-day baseline; fewer days exceeding threshold to compensate for fewer observations
Kassam (4)	Twice per cycle Weekly Twice Weekly	Same as above, but threshold of ratio >4	Higher threshold to compensate for lack of 3 consecutive observations above threshold.
Kassam (5)	Twice per cycle Weekly Twice Weekly	Same as above (ratio >4), but threshold does not have to follow minimum value	Difficult to estimate a good baseline, so lowest days may not occur where generally expected.
Kassam (6)	Every-other-day	Baseline calculated using 3, rather than 5 specimen running average, and threshold of ratio >3 exceeded for 2, not 3, days	Fewer days exceeding threshold to compensate for fewer observations
Baird (1)	Weekly	60% drop in E/P ratio from one sample day to the next sample day	Based on an alternative Baird method (Baird *et al.*, 1991)requiring a 60% decline in the ratio across a 3-day segment. In weekly sampling, this drop must show up from one sample to the next.
Baird (2)	Twice weekly and Every-other-day	60% drop in E/P ratio within 3 consecutive specimens	Same rational as above—this method most closely replicates the original Baird algorithm.
Baird (3)	Twice per cycle	Day 21 E/P 60% less than day 7 E/P	Same rational as above

TABLE II. Cycle lengths in the data sets

Data Set	Mean cycle length (SD) in days	Cycle length range, in days
30 cycles determined ovulatory by ultrasound	27.2 (3.53)	17-37
30 BIMORA cycles visually determined to be ovulatory	26.2 (2.75)	21-33
22 BIMORA cycles visually determined to be anovulatory	33.1 (12.12)	15-66
6 visually indeterminate BIMORA cycles	37.5 (18.94)	23-75

TABLE III. Performance of algorithms for determining the presence or absence of ovulation*

Marker	Sensitivity (%)	Specificity (%)	Misclassification (%)	Misclassification including indeterminate cycles (%)†
Mid-cycle urine LH peak	53	100	27	34
Mid-cycle urine FSH peak	50	100	29	36
Baird DLT algorithm (E1G/PDG ratio)	97	77	12	21
Kassam PDG rise, interval method, ratio threshold ≥ 3	100	100	0	10

*Test *N*=52 menstrual segments (30 ovulatory, 22 anovulatory)
† Test *N*=58 menstrual segments; including 6 visually classified as indeterminate

Table IV. Ovulation day algorithm accuracy (mean marker minus gold [MMGD] standard difference) and precision (MMGD SD, MMGD mode, and percent of estimates falling ±1 and ±2 days of the mode)**.** The "gold standard" is the serum LH peak.

Algorithm	N*	Mean MMGD ± SD	MMGD Mode (% of samples on mode)	% estimates ± 1 day of mode	% estimates ± 2 days of mode
Day of follicle collapse	27	1.04 ± 0.76	1 (85)	96	100
Peak LH serum	27	0.04 ± 0.59	0 (78)	96	100
Peak LH urine	26	0.58 ± 1.14	0 (46)	81	96
Peak FSH urine	16	0.69 ± 1.30	0 (63)	88	94
Peak FSH serum	15	0.13 ± 0.35	0 (87)	100	100
Day of maximum serum E2 concentration	30	-0.73 ± 1.82	-1 (77)	90	93
Day of maximum E1G concentration	30	0.30 ± 2.74	-1 (37)	60	83
Day of maximum E1C concentration	30	0.87 ± 5.99	0 (30)	70	73
DLT Baird (E1C/PDG)	28	0.43 ± 1.14	0 (46)	82	93
DLT Baird (E1G/PDG)	26	0.38 ± 0.85	0 (50)	89	100
First day of PDG rise, interval method	30	3.17 ± 1.60	3 (27)	70	93
First day of PDG rise, cycles method	25	3.48 ± 1.92	4 (28)	56	76
Waller algorithm	30	1.07 ± 1.44	1 (30)	73	90
Combined Hierarchical Method	30	0.60 ± 1.13	0 (50)	90	93

*N= the number of cycles with the marker present, used to calculate the percent of ovulation day estimates falling on or near the mode.

TABLE V. Performance of algorithms for determining the presence or absence of ovulation with simulating intermittent sampling (n= 88 cycles).

Sampling scheme	Method	Sensitivity	Specificity	Misclassification
One sample on day 7 of cycle, one sample on day 21 of cycle	Baird (3)	72%	82%	26%
	Kassam (1)	78%	82%	21%
	Kassam (4) & (5)	68%	82%	28%
Weekly sampling, beginning on day 1 of the cycle	Baird (1)	93%	82%	10%
	Kassam (1)	92%	82%	11%
	Kassam (4)	78%	100%	16%
	Kassam (5)	78%	91%	18%
Weekly sampling beginning on first Monday following start of menses	Baird (1)	92%	82%	11%
	Kassam (1)	93%	77%	11%
	Kassam (4)	90%	95%	9%
	Kassam (5)	92%	95%	7%
Twice-weekly sampling starting on day 1 of the cycle	Baird (2)	100%	64%	10%
	Kassam (1)	95%	55%	16%
	Kassam (2)	95%	64%	13%
	Kassam (3)	97%	91%	5%
	Kassam (4)	92%	82%	11%
	Kassam (5)	97%	68%	11%
Twice-weekly sampling, Mondays and Thursdays, beginning on the first Monday following start of menses	Baird (2)	100%	68%	9%
	Kassam (1)	98%	64%	11%
	Kassam (2)	100%	64%	10%
	Kassam (3)	97%	77%	9%
	Kassam (4)	97%	91%	5%
	Kassam (5)	98%	73%	9%
Every-other-day sampling	Baird (2)	98%	5%	27%
	Kassam (1)	100%	41%	16%
	Kassam (2)	100%	41%	16%
	Kassam (3)	100%	68%	9%
	Kassam (4)	100%	77%	6%
	Kassam (5)	100%	59%	11%
	Kassam (6)	100%	100%	0%

TABLE VI. Precision (MMGD SD) and accuracy (mean MMGD and percentage of cases with estimated ovulation day within ± 2 and 4 days of gold standard) of modified Kassam methods for estimating ovulation day with intermittent sampling (based on n= 30 cycles).

	Weekly sampling, starting on 1st day of menses*	Weekly sampling, starting on 1st Monday of cycle*	Twice-weekly sampling, starting on 1st day of menses†	Twice weekly sampling, Mon & Thurs†	Every-other-day sampling¥
Mean MMGD (MMGD SD)	2.1 (2.7)	0.9 (2.1)	1.6 (2.8)	0.8 (2.0)	-0.4 (2.6)
% ± 2 days of serum LH peak	40%	70%	63%	70%	77%
% ± 4 days of serum LH peak	70%	87%	80%	90%	90%

* day of ovulation is defined as four days before the first sample for which PDG/min PDG is >3.

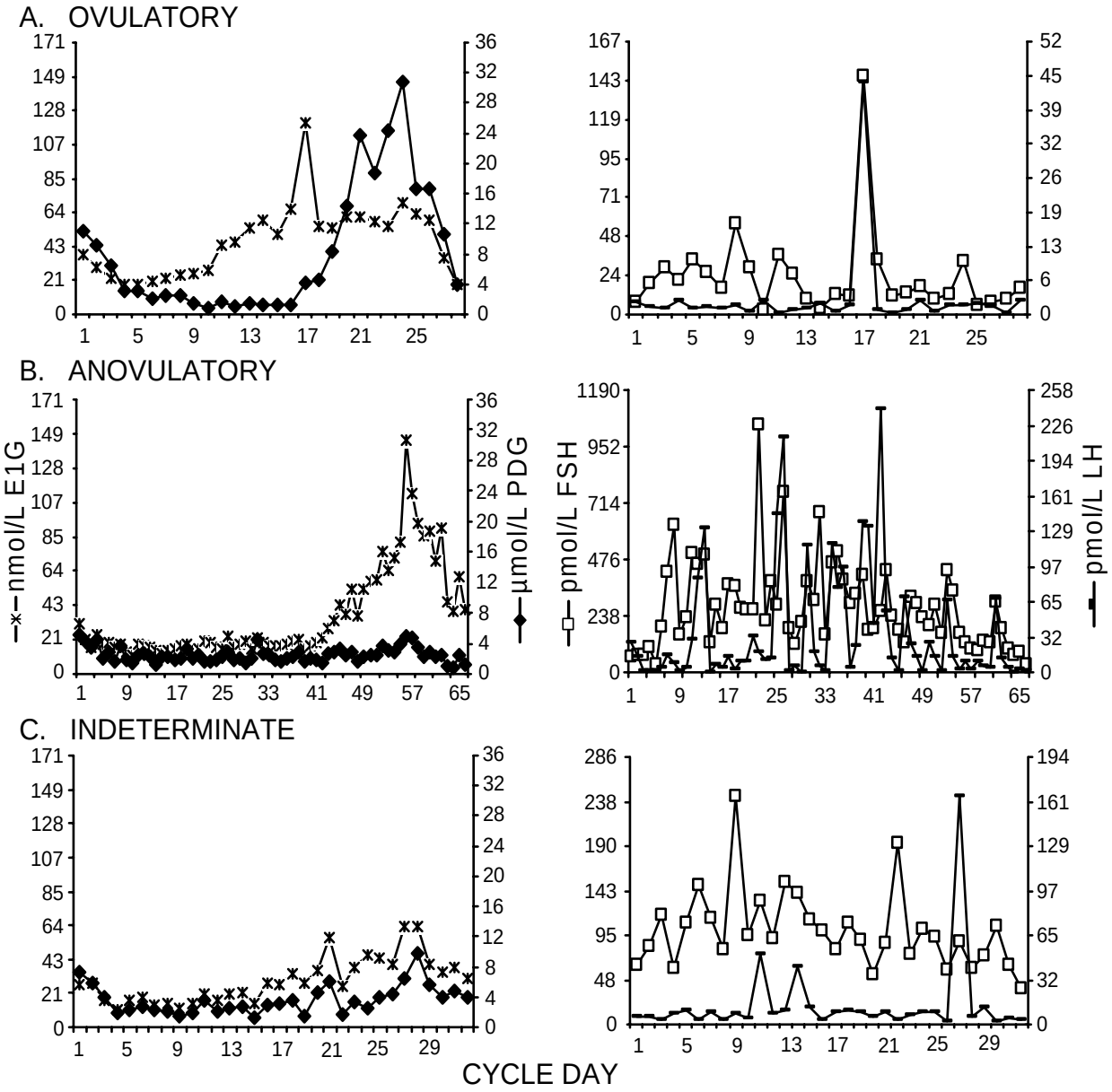
† day of ovulation is defined as on the sample day before the first sample for which PDG/min 3-sample average PDG >3.

¥ day of ovulation is defined as occurring four days before the first of at least two consecutive samples for which the ratio of PDG/min 3-sample average PDG >3.

FIGURE LEGEND

Figure 1. Urinary E1G (*), PDG (♦), LH (—) and FSH (□) profiles of ovulatory (panel A), anovulatory (panel B) and indeterminate (panel C) ovarian cycles.

Figure 1.


A. OVULATORY
B. ANOVULATORY
C. INDETERMINATE
nmol/L E1G
µmol/L PDG
pmol/L FSH
pmol/L LH
CYCLE DAY