Saliva as a Medium for Investigating Intra- and Interindividual Differences in Sex Hormone Levels in Premenopausal Women

Peter H. Gann, Susan Giovanazzi, Linda Van Horn, Amy Branning, and Robert T. Chatterton, Jr.


Abstract
Repeated measurement of ovarian steroids in saliva could provide an advantage in studies estimating long-term sex steroid exposure in premenopausal women, by reducing the measurement error associated with collection of serum or urine samples. We previously reported on characteristics of ultrasensitive RIAs adapted for extraction-free measurement of estradiol (E₂) and progesterone (PG) in saliva. The purpose of the present study was to evaluate the consistency of E₂ and PG levels in saliva in the same women across menstrual cycles, and to compare this with the variation observed between women. We also evaluated the effect of altering the number of consecutive daily samples considered and the method for locating a particular cycle day in relation to ovulation (day 0). Study participants included 12 healthy women who provided daily saliva samples for two consecutive, ovulatory menstrual cycles. A single midluteal serum sample was collected 7–8 days after detection of a luteinizing hormone (LH) peak in urine. We plotted individual cycle profiles and computed intraclass correlation coefficients (ICC) for various definitions of peak and cumulative daily hormone level. For peak PG, determined as the maximal running 3-day mean, ICC was 0.68. For cumulative PG, based on 8 consecutive cycle days (+2 to +9), ICCs were 0.72–0.76 when reverse dating LH peak or rise in salivary PG determined day 0. For E₂, ICCs ranged from 0.74 to 0.79 by various dating methods for the 5 preovulatory days (-4 to 0), and from 0.85 to 0.92 for the 15 days about the center of the cycle (-6 to +8). With exclusion of just the first 5 days of the cycle, the ICC for E₂ was 0.91. For both E₂ and PG, selection of 5 or 7 days for the estimation of the midluteal mean level provided separation of within and between subject variance that was comparable with a LH-timed serum sample. These results indicate that daily saliva samples can be combined to clarify the interindividual differences in E₂ and PG levels in premenopausal women, and that these interindividual differences may be greater than previously imagined.

Introduction
According to the dominant paradigm, a woman’s risk of developing breast (and perhaps endometrial) cancer is affected by long-term exposure to ovarian sex steroid hormones, primarily E₂ and PG (1). Interindividual differences in cumulative exposure to these steroids are, therefore, presumably important. It has proved difficult, however, to test aspects of this model in either etiological or intervention studies involving premenopausal women because of the fluctuation in these hormone levels across the menstrual cycle (2). Previous studies have reported a low correlation between two luteal-phase serum E₂ samples obtained from the same woman over a 1-year interval, although sampling occurred approximately the same number of days after the onset of menses (3, 4). Standardized timing of blood samples can be improved by linking sampling to detection of ovulation or LH peaks, but such testing can be elaborate and/or time-consuming. We have adapted ultrasensitive RIAs for the direct measurement of E₂ and PG in saliva, and have previously reported on the sensitivity, reliability, and serum-saliva correlations for these assays (5). Saliva has several advantages over blood as a sampling medium: it can be easily collected by subjects themselves at repeated intervals; it requires no special collection or storage equipment; and the steroid concentrations measured exclude the fraction tightly bound to serum proteins and thus unavailable for biological action. Most importantly, consecutive daily samples can be grouped for analysis after the length of the menstrual cycle is known.

The aim of the present study was to determine how salivary levels of E₂ and PG track within women from cycle to cycle and to evaluate how cumulative levels of salivary steroid, based on daily sampling, can be used to discriminate interindividual from intra-individual differences. The ultimate goal is to develop a protocol for salivary measurement that reflects sustained exposure of target tissues to ovarian steroids, and thus provides a tool either for etiological studies relating this exposure...
sure to disease risk or for experimental studies of dietary or pharmacological interventions designed to reduce exposure.

Materials and Methods

Study Population. We recruited female participants from the Chicago area with the following eligibility criteria: ages 20–40, no use of exogenous hormones within 6 months, regular menstrual cycles, and at least 12 months since previous childbirth or lactation. Potentially eligible participants were invited to the GCRC at Northwestern University Medical School, where specially trained staff reviewed eligibility, obtained informed consent, and collected baseline data on body size and past medical history. At the initial visit, participants also received supplies and instructions regarding collection of the biological samples during the ensuing two consecutive menstrual cycles. A total of 20 women were recruited, and 19 completed the two-cycle protocol.

Sample Collection. During the two menstrual cycles, each participant provided samples of saliva, urine, breast fluid, and blood. Beginning on the first day after the onset of menstrual bleeding, participants deposited their own saliva, while at home, into sequentially numbered empty plastic vials that were stored in cardboard boxes in the home freezer. The women were instructed to collect the saliva daily between 8 and 11 p.m. Toothbrushing prior to sample collection was forbidden, to reduce the risk of contaminating the saliva sample with blood. After thorough rinsing of the mouth with water, participants chewed sugarless gum to increase salivation while depositing 7–10 ml of saliva into the appropriate vial. The gum was demonstrated previously to have no effect on salivary steroid measurements (6). Saliva collection was continued through the onset of vaginal bleeding marking the end of the second menstrual cycle. Completed boxes of saliva samples were returned to the GCRC, where they were immediately catalogued and stored at −20°C.

For each participant, study staff calculated the expected day of ovulation based on the usual menstrual cycle length. Five days prior to this predicted day, participants began testing their first-void morning urine specimens using the Ovu-Quick cassette for urinary LH (Quidel, San Diego, CA). Previous studies have validated this method for the detection of LH in urine (7). On the day the LH peak was detected, participants noted the day on a calendar provided and called the GCRC to schedule a midluteal appointment for 7 days after the LH peak. Midluteal appointments falling on Sunday were scheduled for the following day. In the event of a failure to detect a clear LH peak, participants were instructed to come to the GCRC 7 or 8 days after the predicted ovulatory day based on the usual cycle length. Venous blood samples were drawn in the morning after an overnight fast. Serum was separated and stored at −70°C.

Laboratory Methods. All of the samples from a given individual were assayed in duplicate in the same run. Unidentifiable quality control samples were inserted in each batch to allow measurement of blinded intra- and interassay CV. Before analysis, frozen saliva was thawed and centrifuged at 1500 × g for 1 h, with the supernatant saved for assay. Salivary PG was measured with a modification of the competitive immunoassay described previously (8). Briefly, this assay uses tritium-labeled PG and an antiserum prepared by one of the authors (R. T. C.) that has known cross-reactivities of 0.5% with 17-hydroxy PG, 2.9% with pregnenolone, 0.95% with corticosterone, 11.4% with 5α-pregnanedione, 1.9% with 5β-pregnanolone, and <0.1% with seven other steroids tested. Standards were prepared in a special gelatinized buffer [0.1 M PBS (pH 7.0), containing 0.015 M Na2S, and 0.1% gelatin]. The volume of sample used was 0.2 ml. Intra-assay and interassay CV were 13.5 and 18%, respectively.

Salivary E2 was measured with a double antibody RIA, also described previously (5). Antiserum and 125I-labeled E2 tracer were obtained from Diagnostic Services Laboratories (Webster, TX). The antiserum has cross-reactivities of 2.4% with estrone, 0.01% with estrone sulfate, 0.21% with 16-keto-E2, 2.6% with E2-3-glucuronide, 0.64% with estriol, and <0.1% with nonphenolic steroids tested. In the assay, the antiserum was diluted to give ~40% binding. Standards were prepared by diluting a methanolic stock solution of E2 with the same gelatin buffer used for PG. A precipitating antibody solution was prepared by titrating the amount of sheep antirabbit gamma globulin required for precipitation of 0.1 ml of rabbit serum, and adding this to propylene glycol (4.8 g/dl). The total volume of sample required was 0.4 ml. Intra- and interassay CV were 9.9 and 11.6%, respectively.

Data Analysis. We assayed salivary PG in all available samples, except those collected in the first 10 days of the cycle, from all of the 19 women who completed both cycles. For each cycle, midcycle day 0 was initially defined in two ways: as the day the urine LH peak was detected or by reverse dating, as the day located 14 days prior to onset of the subsequent menses. After daily salivary PG levels were measured, we identified the day 0 by onset of the rise in PG, defined as the middle day of the first 3-day running mean with an increase of at least 20% over the previous 3-day set, and an average daily level of at least 10 pg/ml. For PG, we defined two consecutive daily segments of interest a priori. The first was the 8-day segment from day +2 to day +9, and the second was the 1-, 3-, 5-, or 7-day segment surrounding the midluteal center. Midluteal center was defined as the midpoint between day 0 and the end of the cycle. These definitions were arbitrary but were chosen to include as many samples as possible while accommodating women with relatively short menstrual cycles. We computed 3-day running means for PG, and we defined peak PG as the middle day of the highest running mean.

If a single daily saliva sample was missing, we imputed a value by averaging levels from the preceding and following days. If two or more consecutive daily samples were missing, or if a given segment contained more than two missing samples, we excluded the segment from the analysis. Of the 19 women whose salivary PG levels were assayed, two had anovulatory cycles (one each) identified by the absence of LH peak and a serum PG of <3 ng/ml. Five other women had at least one cycle with segments disqualified by missing samples. Two women had one cycle apiece in which LH was not detected in the urine, although serum and salivary PG suggested that ovulation had occurred.

For E2, we defined four segments of interest: the 5-day segment from day −4 to day 0 just prior to ovulation, the 15-day segment from day −6 to day +8, the mean daily E2 from all samples excluding the first 5 days, and the midluteal segment encompassing 1, 3, 5, or 7 days. We assayed all of the available saliva samples from 10 women, excluding those with anovulatory cycles or excessive numbers of missing samples. Peak E2 was not determined because a single peak was not expected.

We used repeated measures ANOVA (SAS-PC, SAS Institute, Inc., Cary, NC) to compute between-person and within-person variances. ICCs were computed as the between-person variance divided by the sum of between- and within-person
variances (9). We calculated confidence intervals (95%) for the ICCs based on the expected distribution of the $F$ statistic (9).

**Results**

Among the 12 participants whose cycles were analyzed, mean age was 35 years (range, 26–39 years), 6 were nulliparous, mean body mass index was 23.7, and mean age at menarche was 12.6 years. Mean daily saliva concentrations of E2 and PG across the menstrual cycle, aligned with the urine LH peak as day 0, are shown in Fig. 1. The E2 and PG curves correspond to the daily profiles normally observed in blood but at far lower concentrations and with peaks of smaller amplitude. Day 0 by LH testing corresponded to the day immediately before the onset of the rise in mean PG, and to the day immediately after the steep midcycle drop in mean E2. Fig. 2 shows daily profiles of E2 and PG for consecutive cycles from selected individual participants, aligned by LH peak. These representative curves show greater unexplained day-to-day variation in E2, leading to profiles that contain a discernible preovulatory peak, but few other features typical of a classical serum profile. For PG, however, salivary profiles always contained a distinct rise in PG at the onset of the luteal phase, followed by a relatively smooth, broad luteal peak. Individual and group cycle profiles were similar when reverse dating or rise in PG was used for alignment. Interindividual differences in cumulative levels of both E2 and PG from cycle to cycle are readily apparent.

Cycle-to-cycle variability in peak and cumulative salivary PG is summarized in Table 1. For 24 eligible cycles, the ICC for the peak level of PG was 0.68, meaning that only 32% of the total variance in peak PG was attributable to within-woman variation. For cumulative PG measured over 8 consecutive luteal days (+2 to +9), between-woman differences accounted for three times as much total variance as within-woman differences. These ICCs are all significantly different from the null hypothesis of equal within- and between-woman variance.

We have found that measurement of consecutive daily salivary E2 and PG provides estimates of cumulative and peak concentrations that are consistent within women from one cycle to the next, and are relatively distinct between women. The ICC for peak PG was 0.68 and ranged from 0.72–0.76 for an 8-day luteal segment, depending on the method used for locating ovulation. A single midluteal saliva sample generally gave low ICCs for PG, but increasing the number of consecutive daily samples considered to 3, 5, or 7 produced ICCs that were comparable with that observed with a single serum carefully timed in relation to the urine LH peak. For E2, we observed substantial unexplained day-to-day within-person variation in salivary concentration, but despite this, the combination of consecutive daily saliva samples provided ICCs comparable with, or greater than, that provided by an LH-timed serum sample. When all but the first 5 cycle days were considered, only 9% of the total variance in mean daily salivary E2 was attributable to within-woman variance. This represents much higher within-person correlation than observed for long-established measures used in population studies, such as serum total cholesterol measured 1 year apart ($r$, 0.65; Ref. 10). It appears that the variation in daily salivary E2 coexists with an even greater variation in time-integrated levels between women. Our data indicate that repeated sampling of saliva can be used to reveal potentially important differences in area-under-the-curve or time-integrated exposure of individual pre-
menopausal women to endogenous sex steroids. The importance of even small differences in usual daily exposure, relative to cancer risk, are amplified because such differences are repeated month after month during a woman’s reproductive years (11). Cross-sectional studies have reported substantially higher ovarian steroid hormone profiles in affluent Western women compared with women from nonindustrialized cultures, which provides evidence that the former have a high cumulative exposure to ovarian steroids attributable to higher exposure during a typical cycle as well as to a greater total number of menstrual cycles (12). Within some agrarian cultures such as the Lese of Africa, seasonal variation in ovarian hormone levels suggests that large changes in caloric intake could be an important factor (13). In the U.S., one study involving measurement of estrogen and PG in daily urine samples from a single cycle in 175 women found lower PG levels in women with early menarche and higher body weight and lower estrogen levels in cigarette smokers but no other notable associations with reproductive variables associated with breast cancer risk (14). Regular exercise, whether vigorous or moderate, has also been associated with lower salivary PG profiles (15). Polymorphisms in genes encoding key enzymes involved in sex steroid synthesis or metabolism are also under investigation as possible determinants of interindividual differences in long-term exposure (16). Thus far, however, it appears that most of the interwoman variation in ovarian steroid levels remains unexplained.

Table 1  Variation within versus between women in peak and cumulative salivary PG, measured during two menstrual cycles

<table>
<thead>
<tr>
<th></th>
<th>No. of cycles (subjects)</th>
<th>First cycle mean, pg/ml (s.e.)</th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak PG (3-day running mean)</td>
<td>24 (12)</td>
<td>120.2 (5.8)</td>
<td>0.68</td>
<td>0.22–0.89</td>
</tr>
<tr>
<td>Cumulative PG (days +2 to +9); midcycle (day 0) set by:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse dating⁶</td>
<td>24 (12)</td>
<td>722.6 (55.5)</td>
<td>0.72</td>
<td>0.30–0.91</td>
</tr>
<tr>
<td>Urine LH testing</td>
<td>20 (10)</td>
<td>776.0 (67.8)</td>
<td>0.76</td>
<td>0.32–0.93</td>
</tr>
<tr>
<td>Rise in salivary PG</td>
<td>22 (11)</td>
<td>754.3 (61.8)</td>
<td>0.76</td>
<td>0.37–0.92</td>
</tr>
</tbody>
</table>

⁶CI, confidence interval.
⁶Reverse dating identifies day 0 by subtracting 14 days from the onset of the subsequent menses.

Fig. 2. Individual daily salivary PG and E₂ profiles for pairs of cycles for selected subjects.
We are aware of no previous studies comparing within- and between-woman variability in salivary E\textsubscript{2}. A single study of salivary PG in six cycles from each of eight women reported that interindividual variance was approximately three times the intraindividual variance, a result similar to ours (17). Several studies, however, have examined reproducibility of sex steroid hormone levels in serum or urine samples. Studies involving repeated measures in blood are few and tend to be small because of the obvious practical constraints on performing many serial venipunctures. The largest such study measured daily plasma hormone levels during two cycles from 17 women (18). These investigators calculated the within-person correlation between cycles, roughly equivalent to the ICC, for plasma E\textsubscript{2} and PG during the interval +2 to +8 days after the LH peak. The correlation for PG was 0.80, very close to the estimate of 0.76 that we obtained in saliva for nearly the identical interval. The between-cycle correlation for plasma E\textsubscript{2} was 0.49; although we did not assess this interval in saliva, the plasma result is very comparable with the ICCs that we obtained for a 5-day midluteal segment (0.48–0.60) and is considerably lower than we obtained for longer cycle segments.

In a study involving a single-luteal-phase serum sample, repeated 1 year later, Muti et al. reported an ICC for E\textsubscript{2} of only 0.06 (3). More recently, Michaud et al., using a similar design, obtained an ICC of only 0.19 for a single luteal serum sample (4). However, these investigators timed each blood sample collection to coincide with the anticipated midluteal day based on the individual’s usual cycle length and also had subjects report back on the onset of the subsequent menstrual cycle. They found that with the exclusion of women with probable anovulatory cycles or samples obtained outside a 4–10-day window before the subsequent period, the E\textsubscript{2} ICC increased to 0.62. We note that our results, which also excluded anovulatory cycles, indicate a somewhat higher ICC (0.81) for a single serum sample timed more precisely, 7–8 days after detection of the urine LH peak. The method for timing a single blood sample used by Michaud et al. would have obtained a usable blood sample in 63% of all cycles. Use of saliva might exclude fewer cycles and provide higher ICCs over long segments, but it is not clear that these advantages outweigh the potentially higher costs of collecting daily saliva samples in all conceivable study applications.

Table 2  Variation within versus between women in cumulative and mean daily E\textsubscript{2}, measured during two menstrual cycles

<table>
<thead>
<tr>
<th>Cumulative E\textsubscript{2}</th>
<th>No. of cycles (subjects)</th>
<th>First cycle mean, pg/ml</th>
<th>ICC (SE)</th>
<th>95% CI$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preevulatory (days −4 to 0)</td>
<td>18 (9)</td>
<td>51.5 (7.5)</td>
<td>0.79</td>
<td>0.35–0.95</td>
</tr>
<tr>
<td>Reverse dating</td>
<td>16 (8)</td>
<td>55.1 (7.9)</td>
<td>0.87</td>
<td>0.27–0.95</td>
</tr>
<tr>
<td>Urine LH testing</td>
<td>20 (10)</td>
<td>51.9 (6.8)</td>
<td>0.74</td>
<td>0.27–0.93</td>
</tr>
<tr>
<td>Rise in salivary PG</td>
<td>20 (10)</td>
<td>51.9 (6.8)</td>
<td>0.74</td>
<td>0.27–0.93</td>
</tr>
<tr>
<td>Center cycle (days −6 to +8)</td>
<td>20 (10)</td>
<td>148.8 (19.4)</td>
<td>0.85</td>
<td>0.54–0.96</td>
</tr>
<tr>
<td>Reverse dating</td>
<td>20 (10)</td>
<td>152.5 (23.7)</td>
<td>0.92</td>
<td>0.68–0.98</td>
</tr>
<tr>
<td>Urine LH testing</td>
<td>20 (10)</td>
<td>148.1 (19.3)</td>
<td>0.86</td>
<td>0.56–0.96</td>
</tr>
<tr>
<td>Rise in salivary PG</td>
<td>20 (10)</td>
<td>9.5 (1.4)</td>
<td>0.91</td>
<td>0.68–0.98</td>
</tr>
<tr>
<td>Mean daily E\textsubscript{2}</td>
<td>20 (10)</td>
<td>9.5 (1.4)</td>
<td>0.91</td>
<td>0.68–0.98</td>
</tr>
</tbody>
</table>

$^a$ CI, confidence interval.

Table 3  ICCs for midluteal$^b$ PG and E\textsubscript{2}, measured by a single, LH-timed serum sample versus varying numbers of consecutive saliva samples

<table>
<thead>
<tr>
<th>No. of consecutive daily samples</th>
<th>Saliva PG</th>
<th>Serum progesterone$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reverse dating 0.21 (0.63)</td>
<td>0.77 (0.41–0.92)</td>
</tr>
<tr>
<td>3</td>
<td>Urine LH 0.73 (0.83)</td>
<td>0.73 (0.22–0.93)</td>
</tr>
<tr>
<td>5</td>
<td>Rise in salivary PG 0.16 (0.58)</td>
<td>0.81 (0.42–0.95)</td>
</tr>
<tr>
<td>7 (95% CI)</td>
<td>Serum progesterone$^c$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Urine LH (95% CI)</td>
<td>0.81 (0.46–0.95)</td>
</tr>
</tbody>
</table>

$^b$ Midluteal day determined as midway between day 0 and end of menstrual cycle.

$^c$ CI = confidence interval.

$^d$ Based on single fasting blood sample obtained 7–8 days after urine LH peak.

Apart from the use of sensitive and precise direct assays for salivary hormones, the strengths of this study include estimation of ovulation by urine LH peak and rise in salivary PG as well as by reverse dating. Locating ovulation by detecting the rise in daily salivary PG is an established technique (22), as
is detection of the midcycle decrease in salivary E₂ (23). We found that the rise in salivary PG, the drop in salivary E₂, and reverse dating all gave similar results to LH testing in identifying midcycle day 0. In only one cycle did the selected day 0 differ by more than 3 days between any two methods. Measurement of individual daily saliva samples is relatively expensive. Therefore, in some field settings, it might be more practical to use self-detection of the urine LH peak; and, in fact, our data show that a single serum sample collected at a fixed interval after LH detection is highly reproducible across cycles. However, some women with ovulatory cycles are not able to detect an LH peak with this method, and the daily urine testing around midcycle places a definite burden on study subjects and staff. The ICCs that we observed for location of saliva samples by reverse dating were comparable with those obtained with LH or rise in PG dating. Therefore, we believe that it is most practical to collect samples from an entire cycle and then select samples from a broad consecutive segment located by reverse dating. The required sample volumes for a single assay for E₂ and PG are 400 µl and 200 µl, respectively. It is possible to pool equal aliquots from consecutive days and, thus, perform one assay to measure the cumulative or mean daily level over the segment.

We recognize that our data also have several limitations. The study size was not large, in part because it was necessary to assay many samples per subject in this methodological effort. In addition, we excluded women with anovulatory cycles or excessive missing samples. The criteria for missing samples were quite strict; it is possible that they could be relaxed to include more women and still provide adequate reproducibility, particularly for long cycle segments. Our definitions of cycle segments were somewhat arbitrary. In general, we would expect ICCs to increase as the number of samples that are considered increases. On the other hand, segments that are too long could exclude some women with unusually short cycles. Our study sampled saliva from two consecutive menstrual cycles. It is certainly possible that cumulative levels of sex steroids are more highly correlated in consecutive cycles than those farther apart in time. Individual women do have variations in E₂ from cycle to cycle, and this can even play a role in fecundity during specific cycles (23). However, Lenton et al. (18) reported that the separation between women in plasma PG during the interval +2 to +9 days after the LH peak was easily apparent even when they examined five cycles per woman spanning 3–6 years. Studies ongoing within our group will analyze reproducibility of consecutive daily segments from four cycles occurring over a one-two year period.

In conclusion, these results support the use of salivary measurements for exploring changes in cumulative exposure to endogenous E₂ and PG in premenopausal women. As a recent study has shown, a single midluteal serum sample could be adequate in a large prospective cohort setting in which reverse dating can be used to exclude out-of-phase samples and for which the cost per subject is not great. On the other hand, in trials or small observational studies with a high cost investment per subject, salivary measures could be more efficient by excluding fewer subjects and providing higher ICCs. More methodological research regarding the application of salivary methods is needed. In the context of breast cancer, the relevance of salivary steroid levels to the exposure of breast target tissue itself is a subject of great interest.

Acknowledgments
We thank the staff of the General Clinical Research Center at Northwestern Memorial Hospital for their help in data collection, Kelly Hodges for her excellent work conducting assays, and Laura Studee and Linda Schiffer for their assistance in preparation of the manuscript.

References