BMC Women's Health

Research article

The temporal reliability of serum estrogens, progesterone, gonadotropins, SHBG and urinary estrogen and progesterone metabolites in premenopausal women

Andrew E Williams[1], Gertraud Maskarinec*[1], Adrian A Franke[1] and Frank Z Stanczyk[2]

Address: [1]Cancer Research Center of Hawaii, Honolulu, HI, USA and [2]Department of Obstetrics and Gynecology, and Preventive Medicine, University of Southern California, Keck School of Medicine, Los Angeles, CA, USA

Email: Andrew E Williams - awilliam@crch.hawaii.edu; Gertraud Maskarinec* - gertraud@crch.hawaii.edu; Adrian A Franke - adrian@crch.hawaii.edu; Frank Z Stanczyk - Fstanczyk@socal.rr.com

* Corresponding author †Equal contributors

Published: 23 December 2002

BMC Women's Health 2002, **2**:13

Received: 23 August 2002

Accepted: 23 December 2002

This article is available from: http://www.biomedcentral.com/1472-6874/2/13

Abstract

Background: There is little existing research to guide researchers in estimating the minimum number of measurement occasions required to obtain reliable estimates of serum estrogens, progesterone, gonadotropins, sex hormone-binding globulin (SHBG), and urinary estrogen and progesterone metabolites in premenopausal women.

Methods: Using data from a longitudinal study of 34 women with a mean age of 42.3 years (SD = 2.6), we calculated the minimum number of measurement occasions required to obtain reliable estimates of 12 analytes (8 in blood, 4 in urine). Five samples were obtained over 1 year: at baseline, and after 1, 3, 6, and 12 months. We also calculated the percent of true variance accounted for by a single measurement and intraclass correlation coefficients (*ICC*) between measurement occasions.

Results: Only 2 of the 12 analytes we examined, SHBG and estrone sulfate (E_1S), could be adequately estimated by a single measurement using a minimum reliability standard of having the potential to account for 64% of true variance. Other analytes required from 2 to 12 occasions to account for 81% of the true variance, and 2 to 5 occasions to account for 64% of true variance. *ICCs* ranged from 0.33 for estradiol (E_2) to 0.88 for SHBG. Percent of true variance accounted for by single measurements ranged from 29% for luteinizing hormone (LH) to 92% for SHBG.

Conclusions: Experimental designs that take the natural variability of these analytes into account by obtaining measurements on a sufficient number of occasions will be rewarded with increased power and accuracy.

Background

Several active research programs are investigating the risk associated with serum estrogens, gonadotropins and urinary sex hormone metabolites for a variety of diseases including breast cancer [1], endometrial cancer [2], and osteoporosis [3]. The results of the few published studies suggest that the natural temporal variability (true variation over time, not variation due to storage or other factors) of some serum estrogens, gonadotropins and urinary sex hormone metabolites is sufficiently great that a single measurement occasion may be inadequate to ensure a reliable estimate [4–6]. Published intraclass correlation coefficients (ICC) vary between 0.06 and 0.62 for estradiol (E_2) and between 0.52 and 0.69 for estrone (E_1) [4]. Only the percent of free E_2 and of SHBG-bound E_2 have been found to be sufficiently reliable to account for as much as 50% of the variance in the true mean ($ICC > 0.7$).

The term reliability can refer either to the consistency of a measuring procedure or to the temporal stability of the target of measurement [7]. The definition of temporal reliability used in this study includes both those dimensions, but emphasizes the latter. While researchers can control error due to insufficient repeated measures by increasing the number of measurement occasions, obtaining measurements is expensive. It is therefore useful to have evidence-based guidelines for estimating the minimum number of occasions required to obtain a given degree of reliability for a particular analyte.

All types of measurement error distort, confound, or attenuate the tests of association that constitute one of the primary products of research [8,9]. Figure 1, though not exhaustive, shows the sources of variance in a measurement and the interrelationships between error and tests of model fit or significance.

The relation of a measurement to the object being measured can be represented as: $\sigma_O = \sigma_T + \sigma_E$, where σ_O = variance in the observed measurement of the target, σ_T = variance in the true value of the target, and σ_E = random variance, or error. If the true value of the target is invariant across measurements, i.e., if $\sigma_O = \sigma_E$, the observed variance will be purely a function of the unreliability of the measuring instrument. Conversely, if perfectly error-free measurement of the target could be assumed, i.e., if $\sigma_E = 0$, then $\sigma_O = \sigma_T$ and the observed variance would be purely a function of the temporal stability of the target. If $\sigma_E \neq 0$ and $\sigma_T \neq 0$, the observed variance will be a function of both the temporal stability of the target and of the unreliability of the measuring instrument.

Measurement error can result from a variety of factors, including true variance not captured by a particular measurement strategy, which may complicate the interpretation of temporal reliability estimates. These other factors include variance due to: fluctuations across cycle phases within each woman's menstrual cycle [10]; duration of sample storage prior to analysis [11]; limitations of the assay; multiple analysis batches [10]; multiple types of assays [12]; and multiple laboratories [10]. Ideally, estimates of as many sources of error as possible should be included when considering the impact of temporal reliability on measurement strategy. The objective of this study was to determine the following for various serum estrogens, gonadotropins, and urinary sex hormone metabolites: the minimum number of repeated measurements required for reliable estimates; the *ICC*s; and the amount of true variance accounted for by single measurements.

Methods

Experimental design

The data for this study come from a randomized double-blind study investigating the effects of a 100 mg/day soy isoflavone regimen on estrogen levels in 34 premenopausal women. A detailed description of the study design and the results of the intervention were reported in Maskarinec et al., 2002).)[13]. The Committee on Human Studies at the University of Hawaii approved the study protocol. Written informed consent was obtained from each subject, prior to participation. The study group consisted of 17 premenopausal women per group. Four women left the study before the end of the year and another was able to give only four blood draws for health reasons. Eligibility criteria included: an age range of 35–46 years; an average intake of less than 7 servings of soy foods per week; no prior cancer diagnosis (except basal cell skin carcinoma); no use of oral contraceptives or hormone preparations within the past three months; no intention of becoming pregnant within the next year; an intact uterus and ovaries; self-defined regular menstrual periods; no serious medical condition. Subjects had a mean age of 42.3 years (SD = 2.6), and a mean weight of 65.6 kg (SD = 12.8). Subjects were ethnically diverse: 18 were Caucasian; 6 were Chinese; 5 were Japanese; 5 were Hawaiian.

Sample collection

Subjects were asked to donate 5 urine and blood samples, one at baseline and one after 1, 3, 6, and 12 months of participation. All samples were collected approximately 5 days after the ovulation (approximately day 19 in a 28 day cycle). Subjects used ovulation kits (Ovuquick test kits from Quidel, La Jolla, CA) to determine the time of ovulation. This kit detects the mid-cycle rise of LH using morning urine with a sensitivity of 35 mIU/mL of LH and its predictive validity with respect to ovulation has been estimated as 93% [14]. Although the use of a minimum progesterone value to exclude data from anovulatory cycles from the analyses helped ensure acquisition of the mid-luteal phase samples, only 52% of samples were obtained on exactly the 5th day from ovulation. Ninety-one percent were obtained between the 4th and the 6th day from ovulation. Blood samples were drawn at a commercial laboratory, in the morning between 7 and 9 o'clock to control for circadian rhythm in hormone levels. Serum and urine samples were stored at -80°C after separation and aliquoting.

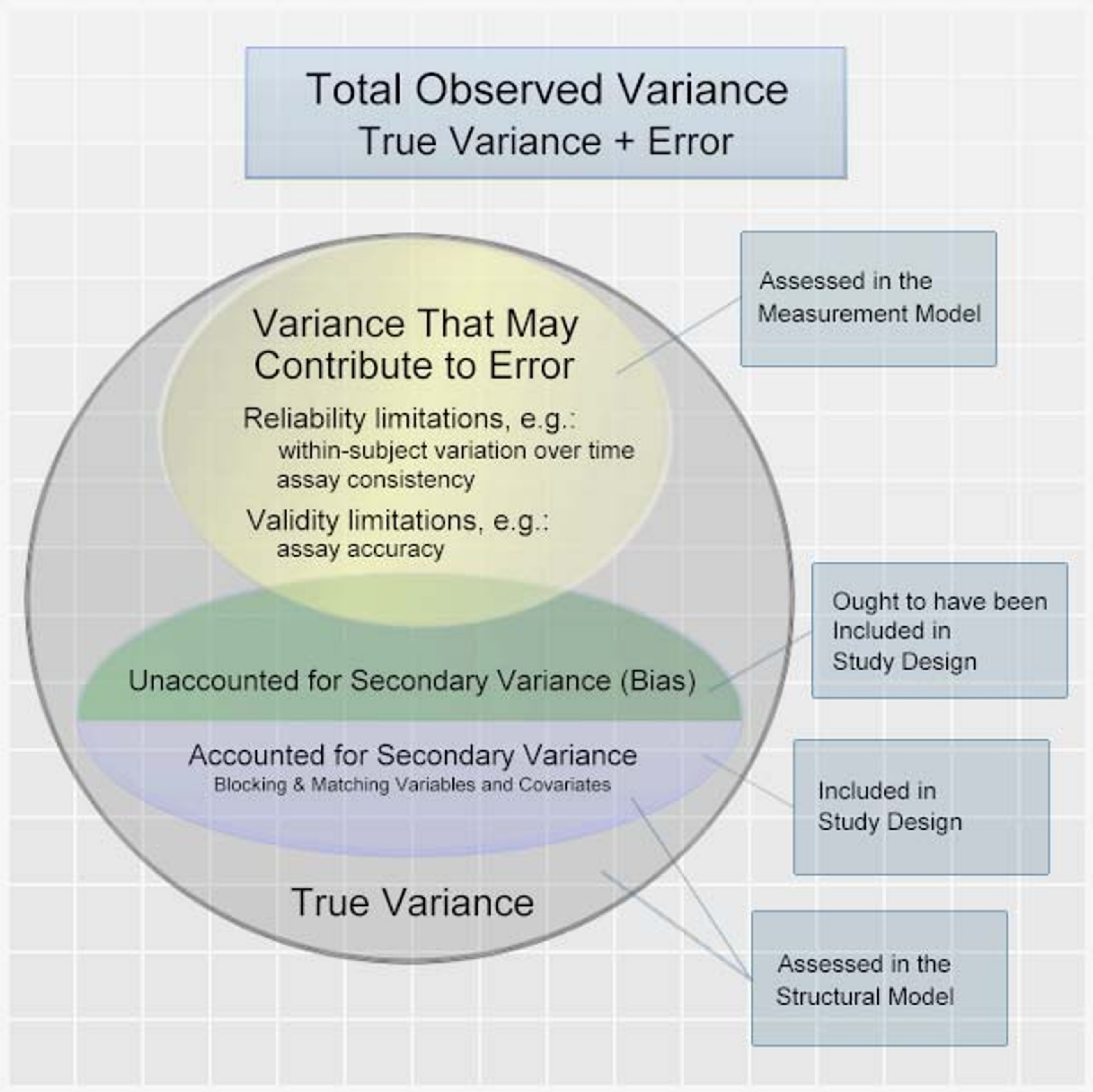
Total Observed Variance
True Variance + Error
Variance That May Contribute to Error
Reliability limitations, e.g.:
within-subject variation over time
assay consistency
Validity limitations, e.g.:
assay accuracy
Unaccounted for Secondary Variance (Bias)
Accounted for Secondary Variance
Blocking & Matching Variables and Covariates
True Variance
Assessed in the Measurement Model
Ought to have been Included in Study Design
Included in Study Design
Assessed in the Structural Model

Figure 1
Total observed variance

Serum analysis

Hormone assays were conducted at the Department of Obstetrics and Gynecology, University of Southern California (Los Angeles, CA) in the Reproductive Endocrine Research Laboratory. The analyses for E_2, free E_2, E_1, E_1S, progesterone, SHBG, follicle stimulating hormone (FSH), and LH were conducted in 2 batches. Samples of these analytes collected at baseline, month 1 and month 3 were analyzed in batch 1, and 6-month and 12-month samples were analyzed in batch 2 one year later. E_2, E_1, progesterone, FSH, LH, and SHBG were quantified in serum by specific and sensitive radioimmunoassays (RIAs). Prior to

Table 1: Coefficients of variation for all analytes

Analyte	Biological Component	Batch 1 Mean of QC Value	CV (%)	Batch 2 Mean of QC Value	CV (%)
Estradiol (pg/mL)	Plasma	53	9.0	38	9.4
		114	7.0	77	9.9
		202	9.0	--	--
Estrone (pg/mL)	Plasma	82	13.0	105	7.8
		158	8.0	262	7.5
		354	7.0	--	--
Estrone-sulfate (ng/mL)*	Plasma	1.1	10.9		
		8.8	11.6		
Progesterone (ng/mL)	Plasma	4.0	7.0	1.1	2.6
		12.0	8.0	9.3	6.1
SHBG (nM/L)*	Plasma	74	2.0		
FSH (mIU/mL)*	Plasma	5.96	7.4		
		16.7	10.4		
		54.2	5.6		
LH (mIU/mL)*	Plasma	1.97	9.1		
		19.2	1.4		
		47.9	1.6		
E_1-G (ng/mL)*	Urine	16.2	7.0		
		33.6	10.3		
PDG (ug/mL)*	Urine	0.6	5.6		
		3.0	2.6		
2-OHE_1 (ng/mL)*	Urine	4.5	2.3		
16α-OHE_1 (ng/mL)*	Urine	2.3	11.6		

*Both batches. QC = quality control. SHBG = Sex hormone-binding globulin. FSH = follicle stimulating hormone. LH = luteinizing hormone. E_1-G = estrone-3-glucuronide. PDG = pregnanediol-3-glucuronide. 2-OHE_1 = 2-hydroxyestrone. 16α-OHE_1 = 16α-hydroxyestrone.

RIA, E_1 and E_2 were first extracted with ethyl acetate: hexane (2:3) and then purified by Celite column partition chromatography, using ethylene glycol as stationary phase [15]. E_1 and E_2 were eluted off the column with 15% and 40% toluene in isooctane, respectively. ^{3}H-E_1 and ^{3}H-E_2 were used as internal standards to follow procedural losses. FSH and LH levels were determined using an immunoradiometric assay (IRMA). E_1S, progesterone and SHBG were measured by direct RIAs using kits obtained from Diagnostic Systems Laboratories, Webster, Texas. Free E_2 (non-SHBG or albumin-bound-E_2) was determined by calculation using a computerized algorithm described previously).)[16]. The majority of intra-assay CVs for all analytes were below 10% (Table 1) indicating good quality control in the laboratory. They ranged from <0.5% for SHBG to 13.0% in the low concentration range of batch 1 for E_1.

Urine analysis

Urine samples were analyzed for estrone-3-glucuronide (E_1-G), pregnanediol-3-glucuronide (PDG), 16α-hydroxyestrone (16α-OHE_1) and 2-hydroxyestrone (2-OHE_1). E_1-G and PDG were measured directly in urine by enzyme immunoassay [17]. Commercially available enzyme-linked immunosorbent assay kits (Estramet: Immuna Care Corporation, Bethlehem, PA) were used to determine levels of 16α-OHE_1 and 2-OHE_1 in urine [18]. All results are relative to creatinine excretion.

Statistical analysis

The SAS statistical software package version 8.2 (SAS Institute Inc., Cary, NC, 1999–2001) was used to perform the statistical analyses. All statistics were computed using logged values when raw values were not normally distributed. To ensure that all measurements in the analysis were from the same time in the menstrual cycle, observations were only included if the concurrent progesterone values were at least 5 ng/mL, a minimum value after an ovulation has occurred. Because analyses for 8 of 12 analytes were conducted in two batches, we included consideration of error due to between batch variance in our analysis of the temporal stability of these analytes. Therefore, estimates of temporal stability for the 8 analytes were calculated for the total number of samples and for the first and second batches separately.

Two types of estimates of the number of measurement occasions (O) necessary to obtain an adequately reliable estimate were computed. The first, the relative type (O_R) includes the between-subject variance. O_R was computed using the formula proposed by Nelson et al. [19]: $O_R = \frac{r^2}{1-r^2} x \frac{S_W^2}{S_B^2}$ where r is the correlation between the observed and the true mean analyte values for an individual over a year, $s_W{}^2$ is the within-subject variance, and $s_B{}^2$ is the between-subject variance. Setting r to 0.9 results in a calculation of the number of measurement occasions required to obtain an estimate that would account for 0.9^2 or 81% of the true variance in the target. Ninety-five percent confidence intervals (95% CI) for O_R were computed using a published method).)[20].

The second estimate of the number of measurement occasions necessary to obtain an adequately reliable estimate, the absolute type (O_A), includes only within-subject variance. O_A was calculated as $O_A = \left(\frac{1.96 \times \sqrt{\sigma_w}}{0.2} \right)^2$, where σ_w is the within-subject variance [21]. By adjusting the denominator, this method allows for the desired approximation to the true mean to be specified as a percentage. Setting the denominator to 0.2 results in a calculation of the number of occasions required to obtain an estimate that is within 20% of the true mean. A SAS macro using Proc Varcomp and Proc Means to produce estimates of O_R, O_A, and related statistics is available from the authors.

*ICC*s measure the proportion of variance attributable to targets of measurement as a ratio of within-subject variance to total variance [22] and are suitable to compare variables of the same measurement class [23]. We computed two types of *ICC*s using the notation developed by Shrout and Fleiss [22]: *ICC(2,1)* was computed for each analyte using all 5 measurement occasions to estimate the temporal reliability of the analyte; *ICC(2,k)* was computed between batches to estimate the contribution of between-batch variance to the temporal reliability estimate. *ICC(2,1)* was computed as *ICC(2,1)* = $\frac{BMS - EMS}{BMS + (k-1)EMS + k(OMS - EMS)}$, where *BMS* is the between-subjects mean square, *EMS* is the error mean square, *k* is the number of observations, *OMS* is the observations mean square, and *n* is the number of subjects [22]. *ICC(2,k)* was computed as $ICC(2,k) = \frac{BMS - EMS}{\left(\frac{BMS + (OMS - EMS)}{n-1} \right)}$. We applied the formulas by Shrout and Fleiss [22] to obtain 95% CIs.

To estimate the percentage of true variance accounted for by a single measurement, we assumed that the best available estimate of the true variance was the total variance for all occasions.

After calculating the Pearson correlation of each occasion with all other occasions, we considered the squared average of these correlations as the estimate of the most likely percent of true variance for which a single occasion could account. We used the formula $\%\sigma_T = \left(\frac{1}{0} \sum_{1-0} r_T \right)^2$, where % σ_T is the percent of true variance, r_T is the Pearson correlation of each occasion with the total of all other occasions, and o is the number of occasions.

Results

Overall means, number of samples, and means by measurement occasion for all analytes (Table 2) indicate the overall stability for the analytes over one year. Although estrogen and progesterone levels were on the average 7% higher and gonadotropins and urinary sex hormone metabolites 10% lower in the intervention than in the control group (data not shown), none of the differences was even close to statistical significance (p values ranged from $p = 0.16$ to $p = 0.90$ for Estrone-sulfate and Estrone respectively). Because of this homogeneity, results in this study were collapsed across experimental groups. The decrease in E_2 and E_1 are the result of laboratory drift and were independent of intervention status).)[13].

The measurement occasions required to obtain a reliable estimate differed considerably by analyte (Table 3). Using the relative method to account for 81% of the true variance, the number of occasions required ranged from O_R = 0.48 to O_R = 11.43 (for SHBG and E_1 respectively). To account for 64% of the true variance, the number of occasions ranged from O_R = 0.20 to O_R = 4.77 (for SHBG and E_1 respectively). Using the absolute method, the number of occasions required to obtain an estimate to within 20% of the true mean, ranged from O_A = 0.34 to O_A = 10.27 (for E_2 and PDG respectively). It appears that, except for SHBG and E_1S, using a single measurement for any of the analytes in this analysis may be problematic for the typical purposes of epidemiological research because the results of typical epidemiological research center on analyses of the mean value obtained from one group *vs.* the mean value obtained from another, e.g. a group of cases or an intervention group *vs.* a control group.

Figures 2 to 4 illustrate the different relationship of between- to within-subject variance and the corresponding difference between O_R and O_A.

Table 2: Basic descriptive data for all measurement occasions of all analytes

	All Measurements		Means for Each Measurement Occasion				
Analyte	N	M (SD)	1 (BL)	2 (1 mo)	3 (3 mo)	4 (6 mo)	5 (1 yr)
Estradiol (pg/mL)	162	133.32 (51.41)	141.44	140.12	151.78	114.70	116.66
Free Estradiol (pg/mL)	159	3.22 (1.21)	3.44	3.34	3.58	2.74	2.95
Estrone (pg/mL)	162	103.74 (34.77)	115.82	115.29	122.59	81.24	80.83
Estrone-sulfate (ng/mL)	160	4.20 (2.44)	4.74	4.81	4.16	3.80	3.41
Progesterone (ng/mL)	162	9.81 (4.89)	10.09	8.87	11.09	10.16	8.76
SHBG (nmol/mL)	159	48.56 (21.18)	49.81	50.01	50.49	46.15	45.73
FSH (mIU/mL)	159	4.78 (4.21)	3.68	4.89	4.77	5.83	4.73
LH (miu/ml)	161	5.12 (4.34)	4.46	6.46	5.56	4.82	4.18
E_1-G (ng/ml)	164	27.74 (20.79)	28.34	31.04	26.06	27.50	25.43
PDG (ug/ml)	164	3.47 (2.15)	3.93	3.23	3.67	3.36	3.11
2-OHE$_1$ (ng/ml)	164	13.76 (8.01)	13.32	14.55	13.07	15.10	12.65
16α-OHE$_1$ (ng/ml)	164	6.94 (5.18)	6.78	6.97	6.42	8.77	5.62

Measurements 1–3 were analyzed in batch 1 and measurements 4 & 5 were analyzed in batch 2. SHBG = Sex hormone-binding globulin. FSH = follicle stimulating hormone. LH = luteinizing hormone. E $_1$-G = estrone-3-glucuronide. PDG = pregnanediol-3-glucuronide. 2-OHE$_1$ = 2-hydroxyestrone. 16α-OHE$_1$ = 16α-hydroxyestrone.

Table 3: Minimum occasions required to obtain a reliable estimate, intraclass correlation coefficients, and percent of true variance accounted for by single measurements

Analyte	Occasions (samples)	Measurement Occasions Required		*ICC(2,1)* (95% CI)	% of True Variance Accounted for by a Single Occasion (Range)
		To Account for 81% of True Variance	To be Within 20% of True Mean		
		Relative Method (95% CI)	Absolute Method		
Estradiol	5 (100)	8.26 (4.53–13.88)	0.34	0.33 (0.18-0.51)	37 (18-74)
Free Estradiol	5 (85)	5.32 (2.92-8.94)	2.00	0.41 (0.26-0.59)	48 (29-72)
Estrone*	3 (60)	11.43 (6.26-19.20)	0.37	0.51 (0.30-0.69)	50 (14-56)
Estrone-sulfate	5 (90)	1.42 (0.78-2.38)	2.03	0.71 (0.58-0.82)	74 (53-77)
Progesterone	5 (100)	5.15 (2.82-8.65)	8.90	0.40 (0.25-0.58)	40 (27-50)
SHBG	5 (85)	0.48 (0.26-0.80)	1.78	0.88 (0.81-0.93)	92 (84-96)
FSH	5 (85)	5.11 (2.80-8.59)	2.63	0.40 (0.25-0.58)	37 (22-66)
E_1-G	5 (100)	3.88 (2.13-6.52)	0.92	0.47 (0.32-0.64)	45 (29-55)
LH	5 (95)	8.38 (4.59-14.07)	8.08	0.30 (0.15-0.48)	29 (17-62)
PDG	5 (100)	5.17 (2.84-8.69)	10.27	0.40 (0.25-0.58)	32 (15-56)
2-OHE$_1$	5 (100)	4.21 (2.31-7.07)	1.57	0.46 (0.30-0.63)	16 (4-40)
16α-OHE$_1$	5 (100)	2.71 (1.48-4.55)	2.44	0.56 (0.41-0.71)	46 (14-79)

Note: includes only complete observations where progesterone > 5 ng/mL; logarithmic transformations of values were used if raw values were not normally distributed. * Calculations based on batch 1 only due to high inter-batch variance. SHBG = Sex hormone-binding globulin. FSH = follicle stimulating hormone. LH = luteinizing hormone. E_1-G = estrone-3-glucuronide. PDG = pregnanediol-3-glucuronide. 2-OHE $_1$ = 2-hydroxyestrone. 16α-OHE $_1$ = 16α-hydroxyestrone.

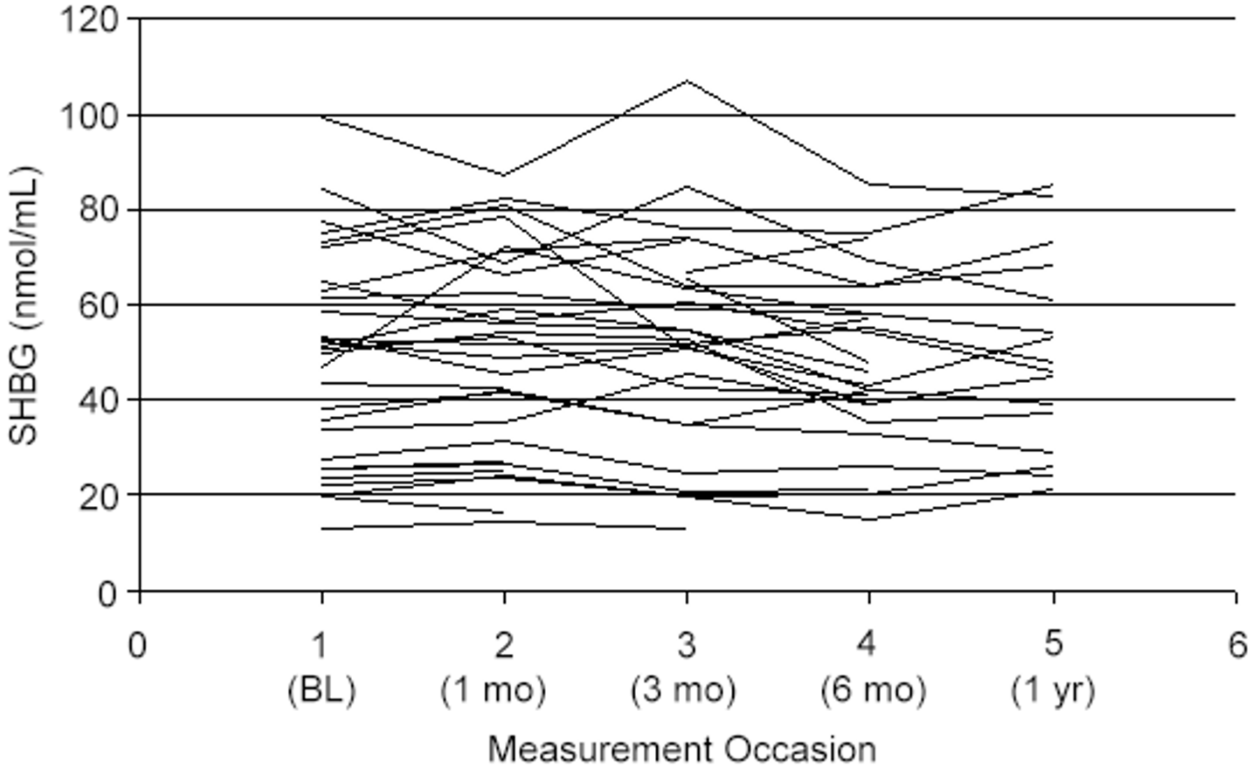

Figure 2
Sex hormone-binding globulin values for all participants by measurement occasion

In the case of SHBG (Figure 2), within-subject variance is small relative to between-subject variance. There is little variation within subjects relative to the variation between subjects, resulting in small O_R and O_A estimates (0.48 and 1.78 respectively). The PDG values (Figure 3) illustrate the case in which within subject variation is high and overlap one another considerably, resulting in relatively large O_R and O_A estimates (5.17 and 10.27 respectively). Finally, Figure 4 depicts the case in which within-subject variance is small, but so is the variance between subjects. In this case, the small within-subject variance results in a small O_A estimate (0.34), but because the within-subject variance is not small relative to the between-subject variance, the O_R is relatively large (8.26).

Because *ICC*s include both within- and between-subject variance, *ICC*s closely followed O_R rather than O_A estimates.*ICC(2,1)* ranged from *ICC(2,1)* = 0.30 to *ICC(2,1)* = 0.88 (for LH and SHBG respectively, Table 3). The intraclass correlation coefficient *ICC*s for absolute agreement between the two analysis batches ranged from *ICC (2,1)* = 0.47 to *ICC (2,1)* = 0.96 (for E_1 and SHBG respectively, Table 4). Estimates of *ICC*s were, generally, consistent across batches, with similar estimates based on analysis of all 5 occasions and for estimates based on each batch. The between batch *ICC* for E_1, however, was less than 0.5, suggesting that the batch 1 *ICC* may be a better indicator than the *ICC* based on all samples. The percent of true variance accounted for by a single measurement ranged from 29% to 92% for LH and SHBG respectively.

Discussion

We have provided estimates to the minimum number of measurement occasions required to ensure adequate reliability for two types of experimental aims. Analyses in epidemiologic studies involve calculations in which between-subject as well as within-subject variance is important. Therefore, O_R will usually be the appropriate index of the minimum number of occasions needed to obtain a reliable estimate. Estimates of O_R based on our sample suggest that only SHBG and E_1S had sufficient temporal stability to be adequately reliable with a single measurement when the desired amount of variance to account for was set as low as 64%. A single measurement of

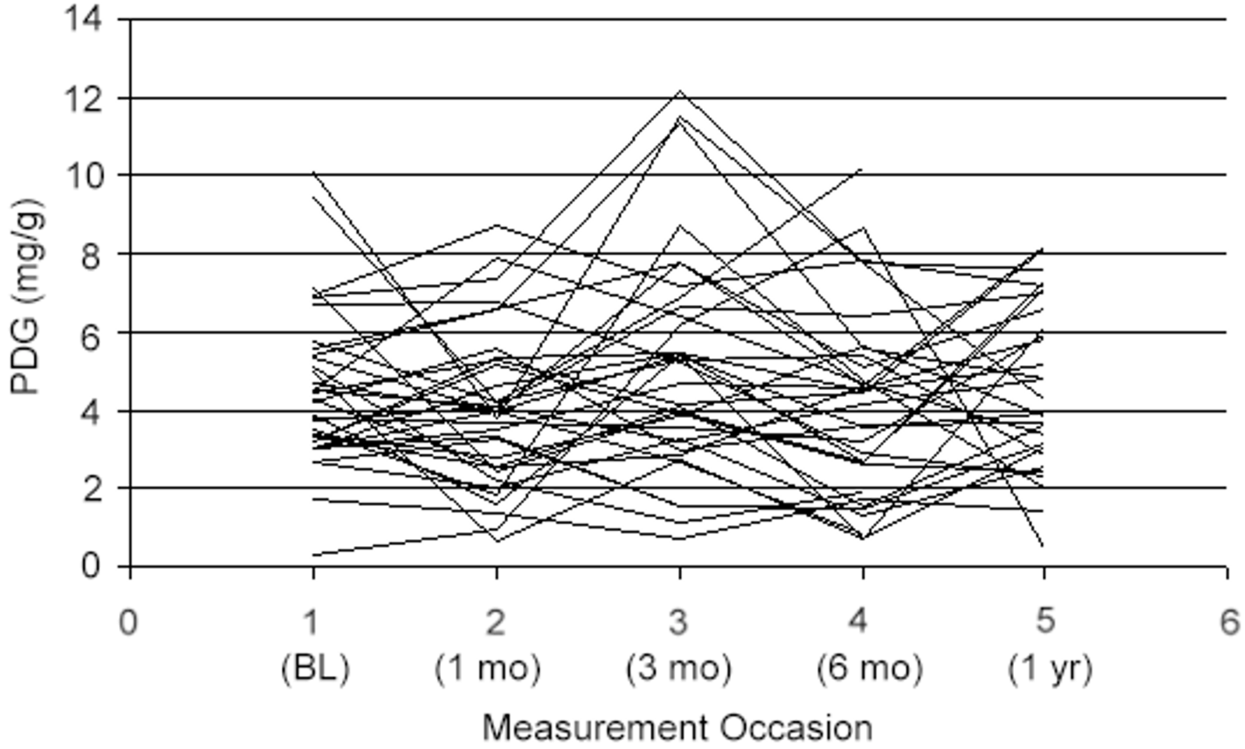

Figure 3
Pregnanediol-3-glucuronide values for all participants by measurement occasion

any of the other analytes would be unlikely to account for even 50% of the true variance. For cases in which the within-subject variance is the only variance of interest, e.g., when the measured value of an analyte will be compared with a fixed standard, O_A will be the appropriate index. The omission of between-subject variance from the formula for calculating this statistic produces very different results from O_R. Several of the analytes that were adequately reliable with a single measurement or very few measurements, when between-subject variance was a factor, required higher numbers of measures when only within-subject variance was involved and vice versa.

This study confirms previous findings that SHBG may be reliably measured in premenopausal women using a single occasion. It also indicates that E_1S may be reliably measured using one sample only. More importantly, our results suggest that none of the other analytes examined meet minimal reliability requirements that would permit confidence in single measures. These results are in agreement with the wide range if *ICCs* reported in previous studies [4–6]. Our conclusions are limited to the collection of samples at midluteal phase, however, and may not generalize to other phases of the menstrual cycle.

The use of *ICC*s to estimate the agreement between analysis batches differs from their use as an index of temporal reliability. The appropriate type of *ICC* for this purpose uses a mean of several values rather than single values and is typically higher than that calculated using single values. Though the *ICC*s between batches were higher than those estimating temporal reliability, they were relatively low, demonstrating the importance of measuring all samples in one batch when possible. As was previously noted [11], error due to time in storage will affect estimates of temporal reliability. Analyzing in multiple batches is one means of decreasing this source of error, but runs the risk of increasing error due to multiple batches. Until better estimates of the impact of storage time on each of these analytes are available, however, it will be difficult to draw conclusions about whether error due to multiple analysis batches or error due to storage time has the more detrimental effect on temporal reliability.

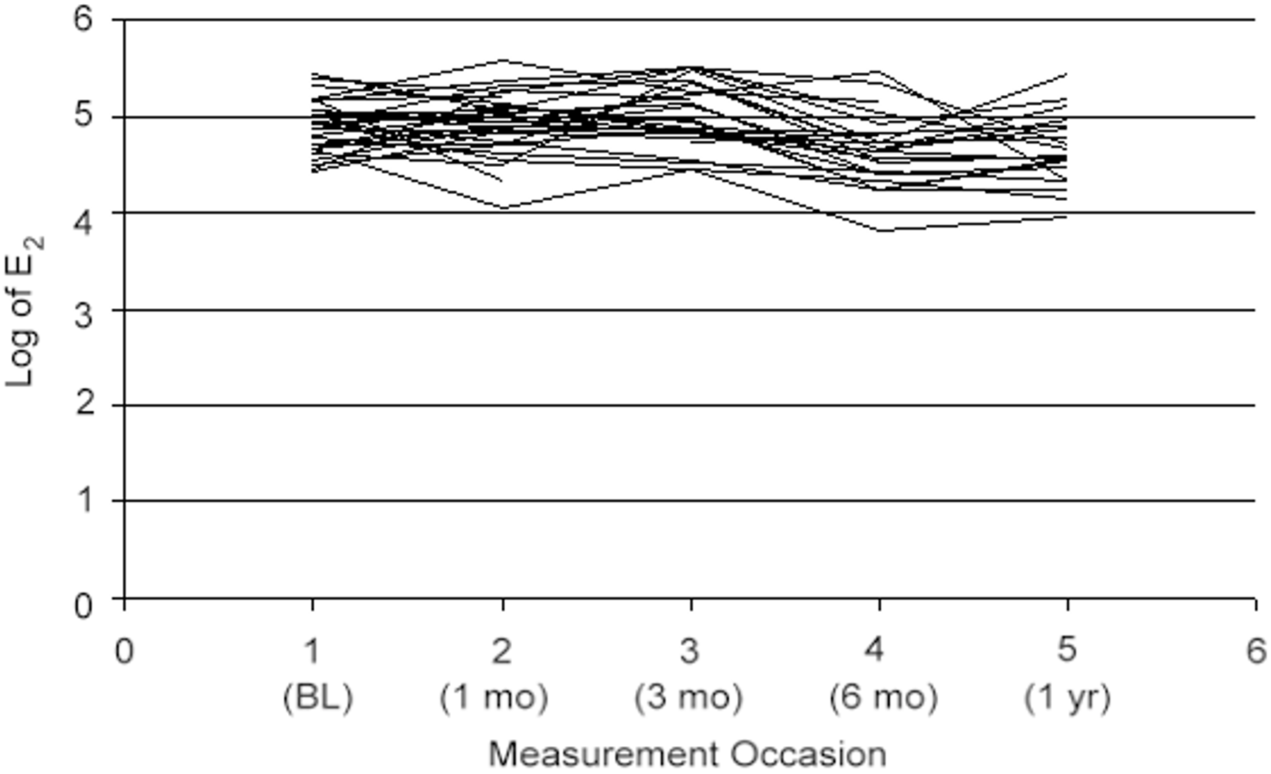

Figure 4
Logged estradiol values for all participants by measurement occasion

Table 4: Intraclass correlation coefficients between batches for analytes analyzed in 2 batches

Analyte	*ICC(2,k)*	95% CI
Estradiol	0.60	-0.03-0.83
Free Estradiol	0.70	0.24-0.87
Estrone	0.47	-0.18-0.79
Estrone-sulfate	0.91	0.77-0.96
Progesterone	0.78	0.58-0.88
SHBG	0.96	0.88-0.99
FSH	0.76	0.57-0.88
LH	0.65	0.36-0.82

Batch 1: 3 occasions over ≈ 3 months. Batch 2: 2 occasions over ≈ 6 months. Total: 5 occasions over ≈ 1 year. SHBG = Sex hormone-binding globulin. FSH = follicle stimulating hormone. LH = luteinizing hormone. E_1-G = estrone-3-glucuronide. PDG = pregnanediol-3-glucuronide. 2-OHE$_1$ = 2-hydroxyestrone. 16α-OHE1 = 16α-hydroxyestrone.

Several sources of error are effectively beyond researchers' capacity to control. For example, the validity and reliability of the best assay available for measuring a given analyte cannot be increased through improving study design. Other sources of error, however, can be dramatically reduced through the use of appropriate designs. These strategies may include, increasing the sample size to reduce the impact of random error, analyzing all samples in one batch, and using a sufficient number of repeated measures to obtain an adequately reliable estimate. It is also possible, though not uncontroversial, to control error statistically by correcting for attenuation using validation data [24].

Several improvements, in addition to a larger sample and more repeated measures, would have increased confidence in the results of our study. First, if the effects of storage time on the analytes were known, we could have taken into account the contributions of this source of variance to our temporal reliability estimates and distinguished its impact from that due to assay reliability. Second, obtaining blood and urine samples on day 5 following ovulation was most appropriate for the measurement of progesterone and near-optimal for SHBG, but may not have been the best day to obtain estimates of the other analytes [25]. Third, though our data were drawn from an intervention study in which no results approached significance, a more clearly homogeneous sample would have been preferable. Fourth, variation in menstrual cycle length and variance due to pulsatility of excretion were additional sources of error.

Finally, our estimates were based on targets that changed across measurements, and we could not assume error-free measurements. Consequently, we were not able to precisely distinguish between the contributions of assay reliability and the contributions of each analyte's natural variability to our estimates of temporal reliability. However, despite some limitations, this study provided significant new insights into the variability of sex hormones, gonadotropins, and urinary hormone metabolites in premenopausal women during a one-year period. Our estimates of temporal reliability represent the combined computation of the consistency of a measure across repeated measurements and the temporal fluctuations in the target of measurement.

Conclusions

Given the relatively large sample size for this analysis and the strictly controlled protocol to collect samples on the same day of the menstrual cycle, our results will be useful for designing future research projects exploring the role of sex hormones in the etiology of cancer and other diseases.

Competing interests

This project was supported by the Pharmavite Corporation in San Fernando, California and a Developmental Funds award from the Cancer Center Support grant to the Cancer Research Center of Hawaii (P30CA071789).

Authors' contributions

AW conceived of the study and performed the statistical analyses. GM was the primary investigator on the original study from which the data for this study was drawn and contributed to the design of this study. FS carried out the immunoassays and contributed to the writing-up of this study. AF participated in the study design and consulted with the authors.

Acknowledgements

The authors gratefully acknowledge the valuable assistance, advice, and guidance provided by Lynne R. Wilkens, Dr PH, and Ian Pagano, MA, both of the Cancer Research Center of Hawaii. We are grateful to the women who donated their time and effort to participate in this study. The project was funded by a contract from the Pharmavite Corporation in San Fernando, California and by a Developmental Funds award from the Cancer Center Support grant to the Cancer Research Center of Hawaii (P30CA071789).

References

1. Muti P, Bradlow HL, Micheli A, Krogh V, Freudenheim JL and Schunemann HJ **Estrogen metabolism and risk of breast cancer: a prospective study of the 2:16alpha-hydroxyestrone ratio in premenopausal and postmenopausal women.** *Epidemiology* 2000, **11**:635-640
2. Parslov M, Lidegaard O, Klintorp S, Pedersen B, Jonsson L and Eriksen PS **Risk factors among young women with endometrial cancer: a Danish case-control study.** *Am J Obstet Gynecol* 2000, **182**:23-29
3. Moreira Kulak CA, Schussheim DH, McMahon DJ, Kurland E, Silverberg SJ and Siris ES **Osteoporosis and low bone mass in premenopausal and perimenopausal women.** *Endocr Pract* 2000, **6**:296-304
4. Michaud DS, Manson JE, Spiegelman D, Barbieri RL, Sepkovic DW and Bradlow HL **Reproducibility of plasma and urinary sex hormone levels in premenopausal women over a one-year period.** *Cancer Epidemiol Biomarkers Prev* 1999, **8**:1059-1064
5. Muti P, Trevisan M, Micheli A, Krogh V, Bolelli G and Sciajno R **Reliability of serum hormones in premenopausal and postmenopausal women over a one-year period.** *Cancer Epidemiol Biomarkers Prev* 1996, **5**:917-922
6. Toniolo P, Koenig KL, Pasternack BS, Banerjee S, Rosenberg C and Shore RE **Reliability of measurements of total, protein-bound, and unbound estradiol in serum.** *Cancer Epidemiol Biomarkers Prev* 1994, **3**:47-50
7. Nunnally JC and Bernstein IH *Psychometric Theory New York: McGraw-Hill, Inc* 1994,
8. Greenland S **Basic methods for sensitivity analysis of biases.** *Int J Epidemiol* 1996, **25**:1107-1116
9. Wong MY, Day NE and Wareham NJ **Measurement error in epidemiology: the design of validation studies II: bivariate situation.** *Stat Med* 1999, **18**:2831-2845
10. Gail MH, Fears TR, Hoover RN, Chandler DW, Donaldson JL and Hyer MB **Reproducibility studies and interlaboratory concordance for assays of serum hormone levels: estrone, estradiol, estrone sulfate, and progesterone.** *Cancer Epidemiol Biomarkers Prev* 1996, **5**:835-844
11. Bolelli G, Muti P, Micheli A, Sciajno R, Franceschetti F and Krogh V **Validity for epidemiological studies of long-term cryoconservation of steroid and protein hormones in serum and plasma.** *Cancer Epidemiol Biomarkers Prev* 1995, **4**:509-513
12. Falk RT, Gail MH, Fears TR, Rossi SC, Stanczyk F and Adlercreutz H **Reproducibility and validity of radioimmunoassays for urinary hormones and metabolites in pre- and postmenopausal women.** *Cancer Epidemiol Biomarkers Prev* 1999, **8**:567-577
13. Maskarinec G, Williams A, Inouye J, Stanczyk F and Franke A **A Randomized isoflavone intervention among premenopausal women.** *Cancer Epidemiol Biomarkers Prev* 2002, **11**:195-201
14. Rudy EB and Estok P **Professional and lay interrater reliability of urinary luteinizing hormone surges measured by Ovu-Quick test.** *J Obstet Gynecol Neonatal Nurs* 1992, **21**:407-411
15. Goebelsmann U, Bernstein GS, Gale JA, Kletzky OA, Nakamura RM and Coulson AH **Serum gonadotropin testosterone estradiol and estrone levels prior to and following bilateral vasectomy.** *In Vasectomy: Immunologic and Pathophysiologic Effects In Animals And Man (Edited by: Lepow IH, Crozier R) New York: Academic Press* 1979, 165

16. Sodergard R, Backstrom T, Shanbag V and Carstensen H **Calculation of free and bound fractions of testosterone and estradiol-17α to human plasma proteins at body temperature.** *Steroid Biochem Mol Biol* 1982, **16:**801-810
17. Munro CJ, Stabenfeldt GH, Cragun JR, Addlego LA, Overstreet JW and Lasley BL **Relationship of serum estradiol and progesterone concentrations to the excretion profiles of their major urinary metabolites as measured by enzyme immunoassay and radioimmunoassay.** *Clin Chem* 1991, **37:**638-644
18. Falk RT, Rossi SC, Fears TR, Sepkovic DW, Migella A and Adlercreutz H **A new ELISA kit for measuring urinary 2-hydroxyestrone, 16alpha-hydroxyestrone, and their ratio: reproducibility, validity, and assay performance after freeze-thaw cycling and preservation by boric acid.** *Cancer Epidemiol Biomarkers Prev* 2000, **9:**81-87
19. Nelson M, Black AE, Morris JA and Cole TJ **Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision.** *American Journal of Clincal Nutrition* 1989, **50:**155-167
20. Wilkens LR, Le Marchand L, Harwood P and Cooney RV **Use of Breath Hydrogen and Methane as Markers of Colonic Fermentation In Epidemiological Studies: Variability in Exretion.** *Cancer Epidemiol Biomarkers Prev* 1994, **3:**149-153
21. Beaton GH, Milner J, Corey P, McGuire V, Cousins M and Stewart E **Sources of variance in 24-hour dietary recall data: Implications for nutrition study desigh and interpretation.** *American Journal of Clinical Nutrition* 1979, **32:**2546-2549
22. Shrout PE and Fleiss JL **Intraclass Correlations: Uses in Assessing Rater Reliability.** *Psychological Bulletin* 1979, **86:**420-428
23. McGraw KO and Wong SP **Forming Inferences About Some Intraclass Correlation Coefficients.** *Psychological Methods* 1996, **1:**30-46
24. Wong MY, Day NE, Bashir SA and Duffy SW **Measurement error in epidemiology: the design of validation studies I: univariate situation.** *Stat Med* 1999, **18:**2815-2829
25. Ahmad N, Pollard M and Unwin N **The optimal timing of blood collection during the menstrual cycle for the assessment of endogenous sex hormones: can interindividual differences in levels over the whole cycle be assessed on a single day?** *Cancer Epidemiol Biomarkers Prev* 2002, **11:**147-151

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1472-6874/2/13/prepub