AMERICAN JOURNAL OF HUMAN BIOLOGY 20:35–42 (2008)

Original Research Article

Interpopulation, Interindividual, Intercycle, and Intracycle Natural Variation in Progesterone Levels: A Quantitative Assessment and Implications for Population Studies

GRAZYNA JASIENSKA[1,2*] AND MICHAL JASIENSKI[3]
[1]*Department of Epidemiology and Population Studies, Collegium Medicum, Jagiellonian University, 31-531 Kraków, Poland*
[2]*Radcliffe Institute for Advanced Study, Harvard University, Cambridge, Massachusetts 02138*
[3]*Nowy Sacz Business School–National-Louis University, 33-300 Nowy Sacz, Poland*

ABSTRACT Methodological challenges in studying sex steroid hormones in premenopausal women result from the existence of variation at three levels: among women from the same population, among menstrual cycles recorded for women at different times of the year, and among days of the same cycle. We partitioned, for a Polish rural population, the natural, nonpathological, variation in salivary progesterone concentrations (measured during 14 days of the luteal phase) into the intracycle component (which accounts for 65% of the total variation) and the among-cycle component (the remaining 35% of the total variation). Out of the among-cycle variation in salivary progesterone, as much as 46% is expressed as differences among individual women (interindividual component); the remaining 54% of variation is due to differences among cycles of individual women (intercycle, within-women component). Such intercycle variation is probably caused by a seasonality of agricultural workload and is much higher than in nonseasonal, industrial populations. We also used bootstrap analyses to generate heuristic recommendations for choosing sample sizes of the number of subjects, number of cycles per woman, and number of days per cycle. Studies in populations with seasonal lifestyles should rely on measurements of at least three cycles per woman. Given the substantial intracycle amplitude in hormone levels to reliably assess biologically and medically relevant variation in ovarian function, at least 7–8 days/cycle should be measured. Am. J. Hum. Biol. 20:35–42, 2008.

The assessment of sex steroid hormones in women is of crucial importance for many areas of human biology, behavior, reproductive health, and disease prevention. Levels of estradiol and progesterone are used to test hypotheses about evolution of female reproduction (Ellison, 2001; Jasienska, 2003), reproductive and behavioral ecology (Ellison, 2003b; Holman et al., 2004; Jasienska and Ellison, 1998, 2004; Vitzthum et al., 1994), cognition, and mate choice (DeBruine et al., 2005; Feinberg et al., 2006; Jones et al., 2005; Mead and Hampson, 1997; Pawlowski and Jasienska, 2005; Schultheiss et al., 2003; Sherwin, 2005; Williams, 1998). They are also focus of research as potential biomarkers of reproductive and systemic aging (Ferrell et al., 2005), risk of osteoporosis (Hillard and Nelson, 2003), and hormone-dependent cancers (Kaaks et al., 2005; Noh et al., 2006; Pike et al., 1993).

However, population and clinical research on sex steroids faces challenges because of the existence of substantial variation in physiological parameters determining hormone concentrations (see also Himmelstein et al., 1990), both among individuals and within individuals. Variation in hormonal levels exists among populations, among women from one population, among different menstrual cycles of the same woman, and finally during menstrual cycles. First two types of variation are due to genetic, developmental, and adult lifestyle factors. For example, age-related variation results from the lowest steroid hormone levels characterizing women several years after the menarche and several years before the menopause, and highest levels between 25 and 35 years of age (Lipson and Ellison, 1992). However, even when cycles of women of similar age are compared often substantial differences in hormonal levels are apparent.

We hypothesize that populations should differ in the amount of exhibited variation in levels of ovarian steroid hormones, because the amount of expected variation should vary in relation to the lifestyles characterizing women in different populations. Nonindustrial populations experience pronounced seasonal changes in lifestyle (e.g., energy intake and energy expenditure) associated, for example, with agricultural workloads (Jasienska and Ellison, 1998; Jasienska and Ellison, 2004; Panter-Brick and Ellison, 1994). Consequently, such populations should be characterized by higher intercycle variation than interindividual variation. In contrast, urban populations, without seasonal changes in lifestyle, should exhibit higher interindividual variation than intercycle variation. Relatively low intercycle variation in progesterone levels among cycles of individual woman have been reported by the previous studies (Gann et al., 2001; Lenton et al., 1983; Sukalich et al., 1994) of urban women from the United States and the United Kingdom.

Our study is the first to investigate, within a traditional rural population, the partitioning of natural, nonpathological, variation in progesterone levels. We have studied the

Grant sponsor: Department of Anthropology, Harvard University, Polish Committee for Scientific Research, Center for Human and Primate Reproductive Ecology (CHaPRE).

*Correspondence to: Grazyna Jasienska, Department of Epidemiology and Population Studies, Jagiellonian University, Grzegórzecka 20, 31-531 Kraków, Poland. E-mail: jasienska@post.harvard.edu

Received 17 October 2006; Revision received 23 March 2007; Accepted 24 March 2007

DOI 10.1002/ajhb.20686

Published online 26 October 2007 in Wiley InterScience (www.interscience.wiley.com).

*TABLE 1. Characteristics of study subjects (*n *= 22) at the beginning of the study*

	Mean	SE
Age	31.2	1.01
Age at menarche	14.4	0.15
Age at first childbirth	22.0	0.66
Number of children[a]	2.6	0.3
Body height (cm)	161.5	0.88
Body weight (kg)	63.4	1.84
Body mass index (kg/m^2)	24.3	0.63
Body fat (%)	27.6	1.37
Length of menstrual cycle[b]	26.8	0.27

[a]Parous women, $n = 19$.
[b]Mean length calculated from 3 to 6 cycles per woman.

extent of variation in salivary levels of progesterone at three levels: among women from the same population, among menstrual cycles recorded for women at different times of the year, and among days of the same cycle. We have also used bootstrap resampling methods to generate heuristically useful recommendations aimed at improving statistical quality of data generated in the area of reproductive endocrinology, human evolutionary ecology, and epidemiology.

MATERIALS AND METHODS

Subjects

Study participants were 22 women from a small agricultural village located in Beskid Wyspowy mountain range in Southern Poland (the Mogielica Human Ecology Study Site). The study was approved by the Committee on the Use of Human Subjects at Harvard University. Women were between 23 and 39 years of age, and met the following criteria for participation: regular menstrual cycles and no fertility problems, no gynecological and chronic disorders (i.e., diabetes and hypo/hyperthyroidism), no instances of taking any hormonal medication or using hormonal contraception during the 6 months before recruitment, and not having been pregnant or lactating during the 6 months before recruitment. Anthropometric characteristics of participants are presented in Table 1. Life in the village is characterized by intense seasonality in physical workload imposed by requirements of haying and harvest season. Levels of energy expenditure of the summer months were significantly higher than the energy expenditure of winter months (Jasienska and Ellison, 2004). Diet was sufficient through the year and women did not loose weight or body fat when physical work was the most intense (Jasienska and Ellison, 1998, 2004).

Data collection

Subjects collected daily saliva samples for a total of 6 months, from June to October 1992 and in January and February 1993. Each woman was provided with a set of polystyrene collection tubes pretreated with sodium azide as a preservative, a calendar for keeping records of sample collection and marking menstrual dates, and pretested chewing gum to be used as the stimulant of saliva flow. Subjects were requested to collect samples daily, in the evening, at least 30 min after last meal. Very few omissions occurred. Samples were stored at room temperature until the end of each collection period and then transported to the laboratory and frozen at -20°C until assayed. It should be noted that preservation of samples with sodium azide is a recommended method for the radioimmunoassay (RIA), but should not be used if samples are analyzed by the enzyme-immunoassay.

Samples belonging to each menstrual cycle were arranged in order starting from the first day of the menstrual bleeding. The day before the onset of the next menstruation was identified as day -1, with the previous days identified correspondingly. The last 18 daily samples of each cycle (days -18 to -1) were assayed for progesterone. The total of 115 menstrual cycles was assayed, with less than 10% daily samples missing due to missed or improper collection or loss during the laboratory procedure. Samples belonging to a particular cycle were analyzed in the same assay, with cycles from two different subjects run in each assay.

Progesterone was measured in each subject's samples by the RIA according to published protocols (Ellison and Lager, 1986). Quality control was maintained through monitoring values of saliva pools at low (follicular), medium (luteal), and high (pregnancy) levels. Assay sensitivity, that is, the smallest amount distinguishable from 0 with 95% confidence, averaged 22.5 pmol/l. Intraassay variability (CV) at the 50% binding point of the standard curve was 6.3%. Interassay variability estimated from pools containing various levels of progesterone averaged 20.2% for low (late follicular/early luteal) pools, 10.7% for medium (mid-luteal) pools, and 13.9% for high (pregnancy) pools.

Statistical analysis of variance components

Data on progesterone concentrations from the last 14 days (-14 to -1) representing the luteal phase of the menstrual cycle were transformed to natural logarithms and analyzed in a two-level nested (hierarchical) analysis of variance (Table 2). In the model, individual women represented the upper level and cycles collected in different months represented the lower level (with cycles nested within women; see Table 3). A random model was used, implying that both individual women and, for the purpose of this article, also individual cycles-months were considered random samples taken from a larger population. Tests of significance followed the traditional (Expected Mean Squares) ANOVA approach, while variance components and their 95% confidence intervals were computed according to the REML (Restricted Maximum Likelihood) procedure, as implemented in the JMP package (Version 5.0, SAS Institute, 2002).

The variance components associated with variation among women (interindividual) and variation among cycles (intercycle) within women were expressed either as fractions of the total variance (which included the residual, variance component) or were expressed relative to each other to allow comparisons with other published studies (Gann et al., 2001; Sukalich et al., 1994). A similar analysis was performed for the mid-luteal phase data (days -11 to -6, which were chosen based on preliminary investigations of within-cycle patterns of variation; Jasienska, unpublished).

Analyses of samples sizes

All bootstrap procedures were written in the Resampling Stats environment (version 4.0 for the Macintosh)

TABLE 2. Two-level random model nested analysis of variance of log-transformed progesterone levels, measured in 22 women from the Mogielica Human Ecology Study Site, Poland

Source	df	SS	*F*	*P*	Variance component	95% Confidence limits	Percentage of total variance	Percentage of two levels
Days −14 to −1 (entire luteal phase)[a]								
Interindividual	21	122.663	4.362	<0.0001	0.072	0.036–0.204	16.5	46
Intercycle (within women)	93	125.210	4.775	<0.0001	0.083	0.059–0.126	19.0	54
Residual	1,396	393.585			0.282		64.5	
Days −11 to −6 (mid-luteal phase)[b]								
Interindividual	21	75.484	5.601	<0.0001	0.108	0.057–0.284	30.9	55
Intercycle (within women)	93	59.904	4.157	<0.0001	0.087	0.060–0.138	24.8	45
Residual	546	84.609			0.155		44.3	

[a]Data from all 14 days of the luteal phase of the cycle.
[b]Data from the mid-luteal phase (6 days).

TABLE 3. Mean daily luteal phase progesterone levels (in pmol/L; computed from untransformed data) during the 6 months of the study

Reverse cycle day	July	August	September	October	January	February
−14	132.3 (70.23, 18)	122.5 (69.94, 20)	170.5 (130.40, 22)	161.4 (82.60, 15)	136.3 (73.75, 18)	120.0 (49.45, 17)
−13	132.9 (76.13, 18)	166.0 (107.45, 20)	182.0 (92.45, 21)	179.6 (74.11, 15)	172.0 (106.75, 20)	142.4 (93.28, 16)
−12	159.0 (64.06, 18)	189.7 (114.75, 19)	225.5 (135.07, 22)	244.3 (137.92, 14)	186.0 (99.01, 19)	205.3 (112.09, 17)
−11	224.7 (122.77, 16)	216.4 (107.39, 19)	275.8 (126.81, 21)	274.7 (150.84, 15)	237.7 (127.41, 18)	212.5 (89.94, 17)
−10	216.7 (92.52, 19)	232.7 (121.30, 21)	273.3 (233.86, 22)	294.7 (104.93, 16)	287.1 (158.33, 20)	287.5 (139.89, 16)
−9	252.8 (144.24, 18)	245.5 (150.46, 20)	265.9 (137.49, 22)	330.0 (123.65, 15)	253.6 (95.47, 19)	297.4 (143.57, 16)
−8	225.6 (120.22, 18)	248.2 (130.36, 21)	231.2 (117.08, 21)	321.9 (167.14, 16)	284.4 (144.99, 19)	298.1 (173.12, 16)
−7	188.5 (106.14, 18)	231.3 (126.97, 20)	251.3 (133.40, 21)	236.9 (137.98, 15)	264.4 (125.82, 20)	252.0 (137.74, 17)
−6	204.4 (108.50, 18)	180.0 (106.17, 18)	226.0 (114.16, 21)	292.3 (183.06, 15)	224.2 (100.48, 20)	211.9 (116.45, 17)
−5	171.5 (85.78, 19)	175.1 (75.54, 21)	193.7 (97.39, 22)	233.9 (108.46, 16)	206.0 (78.40, 20)	183.5 (84.18, 16)
−4	164.1 (100.23, 19)	161.1 (77.06, 19)	182.2 (78.04, 22)	240.5 (167.86, 15)	184.1 (74.61, 18)	144.7 (71.27, 17)
−3	151.2 (72.35, 19)	143.7 (80.36, 17)	159.6 (91.32, 22)	178.6 (118.46, 13)	129.3 (63.21, 18)	135.2 (69.65, 14)
−2	149.2 (74.16, 19)	117.4 (53.54, 21)	101.2 (55.44, 21)	160.7 (134.63, 13)	102.3 (27.96, 18)	115.9 (68.27, 13)
−1	119.1 (45.69, 15)	91.4 (41.70, 15)	107.1 (53.32, 19)	99.4 (68.15, 14)	100.6 (33.22, 15)	86.3 (47.92, 13)

Day −1 represents the last day of the cycle.
Standard deviation and sample size are in parentheses.

and had a similar structure: progesterone data were used either as raw measurements (for the assessment of the optimal number of days per cycle) or as cycle means (for the assessment of the number of women and the number of cycles per woman). The criterion for the assessment of the amount of noise in the results was the coefficient of variation (CV), which quantified the interindividual (among-women) variance in the bootstrapped sample (of a given size). There were 100,000 bootstrap samples taken and average within-sample CV was computed for each sample size. All CVs are presented in figures recalculated relative to the value which always corresponded to the largest tested sample size (of either 6 cycles, 14 days per cycle, or 150 subjects). Additional graphs illustrate the rate with which increasing sample size affects the statistical precision of the estimates.

Interindividual variation (among subjects) was evaluated for different sample sizes (from 2 to 150 subjects) using data from a separate database on 185 Polish women (49 rural and 136 urban) whose hormonal profiles were measured (one cycle per woman). Average concentrations of progesterone during the measured cycles ranged among these women from 20.1 to 368.6 pmol/l (mean 128.18 pmol/l, CV = 0.477). Details and laboratory procedures have been published elsewhere (Jasienska et al., 2004, 2006b). Each woman in the bootstrap analysis was represented by the mean progesterone level of her cycle.

To evaluate the reliability of an approach, encountered in other studies (Gann et al., 2001; Lenton et al., 1983) that estimated intra- versus interindividual variation based on only two cycles per woman, we performed exploratory nested analyses of variance and computed variance components for all possible pairs of cycle-months in our study (15 pair-combinations of six cycles). We also bootstrapped data on the 12 women for whom the data were complete for all six cycles to study how the number of cycles per woman influences the reliability of the estimates of interindividual variation. Each cycle in the bootstrap analysis was represented by its average progesterone level.

The impact of the number of measured days per cycle was investigated using data on 109 Polish women (a subset of 185 subjects; see above) whose luteal progesterone profiles for days −14 to −1 were without missing data points. Altogether 26 rural and 83 urban women were included in bootstrap analyses, generating samples of 1 through 14 days per cycle. Analyses of the urban women used three sets of 40 women randomly chosen from all 83 women; the results for three urban sets and the rural set were averaged.

The amplitude of hormone concentrations during a cycle may be expressed as a ratio of the nontransformed maximum to the minimum value. Although analysis of ratios may have disadvantages (Jasienski and Bazzaz, 1999), their use here is not for statistical inference but solely for the purpose of illustrating the extent of within-cycle variation. In addition, to reduce the impact of the extremely low values from the beginning and the end of the luteal phase

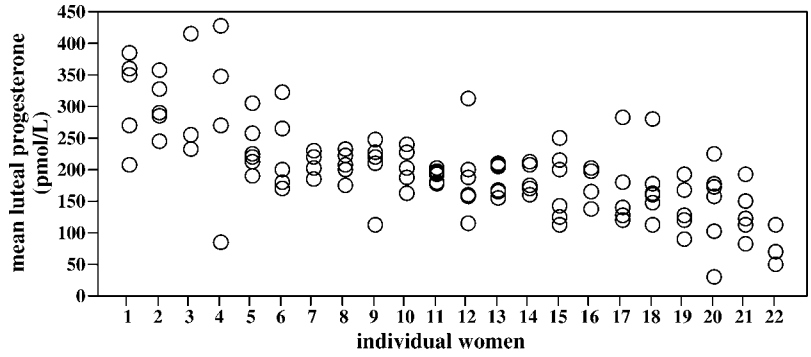

Fig. 1. Interindividual variation among 22 women (ranked according to the mean progesterone level of each woman) and intercycle variation (among cycles of each woman). Each dot represents mean value (computed from untransformed data) of a single cycle of the luteal-phase (days −14 to −1) progesterone level.

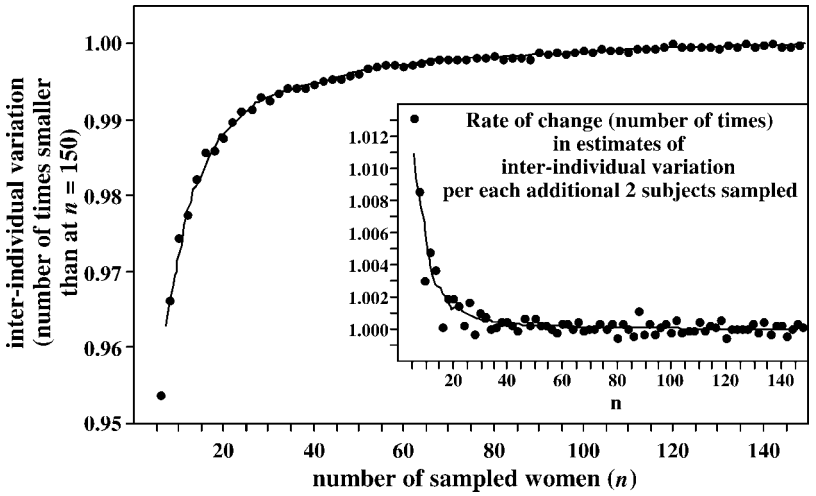

Fig. 2. Interindividual variation increases asymptotically as a function of the number of subjects in the study. The inset shows that estimates of interindividual variation change initially with each additional 2 women sampled, but sampling more subjects than $n = 20$ yields only marginal changes in the estimates of interindividual variation.

(the use of which could result in artifactually high values of amplitude), we also computed amplitudes only for days −12 through −3 of the luteal phase (one cycle had a minimum concentration of zero on one of the days and was excluded).

RESULTS

Partitioning the interindividual and intercycle sources of variation

Two sources of natural variation in progesterone concentrations, i.e., differences among women (interindividual) and differences among cycles of individual women (intercycle), had each highly statistically significant and very similar contributions to the overall observed variation in the data. Expressed relative to each other, i.e., without taking into account the residual (intracycle) variation, the interindividual level accounted for 46%, and the intercycle (within-women) level accounted for 54% of variation, with overlapping 95% confidence limits for variance components (Figs. 1 and 6D; Table 2). When the analysis is performed only on the data from the mid-luteal phase (days −11 to −6), the interindividual and intercycle components of variance become, 55 and 45%, respectively (Fig. 6D; Table 2).

Number of subjects for the assessment of the interindividual variation

If there is only one cycle profile reliably evaluated for each woman, the average CV among 20 subjects' progesterone means is about 99% of that calculated among $n = 150$ subjects, while a sample of six women reaches 95% (Fig. 2). However, if cycle means are based on incomplete hormonal profiles (see below), this fact affects the necessary sample size for the number of subjects. Further, with each subject measured, the interindividual variability (as a measure of both biological variation and sampling variation) increases at a declining rate. Beyond the sample size of 20–25 subjects, each additional two women measured contributed a similar fraction to the precision of the analysis. In other words, the gains from increasing sample size are substantial for samples of up to 20 women.

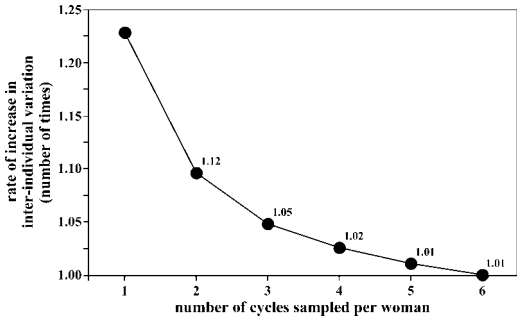

Fig. 3. Interindividual variation as a function of the number of cycles sampled per each subject in the study. The number next to each symbol shows how quickly interindividual variation decreases with each additional cycle measured per woman, e.g. measuring 2 rather than 1 cycles reduces interindividual variation 1.12 times, and measuring 3 rather than 2 cycles reduces interindividual variation an additional 1.05 times. The analysis assumes that cycle means are evaluated reliably (minimum 7-8 days per cycle).

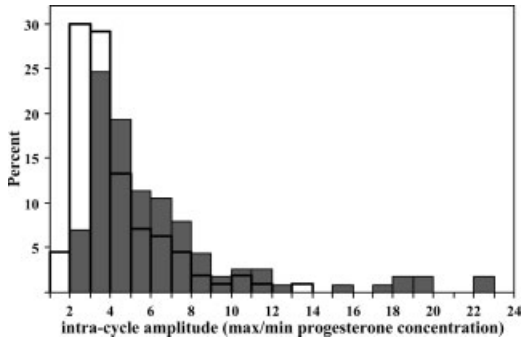

Fig. 4. Distribution of intracycle amplitudes of progesterone concentrations, computed as ratios of untransformed maximum and minimum values recorded between days −1 and −14 (filled bars) and between days −3 and −12 (open bars); $n = 114$ cycles of 22 women recorded during 6 months.

Number of cycles per woman for the assessment of the intercycle variation

Fifteen analyses of variance based on only two cycle-months per woman yielded estimates of interindividual variation (traditionally quantified by the intraclass correlation coefficient, ICC) ranging from 0.04 to 0.87 (mean 0.335). The estimates of ICC varied therefore widely (20-fold) and depended strongly on the choice of particular pairs of cycle-months, showing clearly that a sample of two cycles per woman is not sufficient to capture natural variation among cycles (due in part to a seasonal lifestyle in the studied population).

A bootstrap approach showed that an optimum number of months to be sampled per woman in a seasonal rural population appear to be not fewer than four (Fig. 3). An expected positive covariance across cycles of a single woman (who may tend to produce generally high or generally low levels of hormones) reduces the estimates of intercycle variation and, therefore, means that fewer cycles per woman should be sampled, than without the intercycle covariance. However, a pronounced seasonality of workload or dietary intake will probably reduce such intraindividual, intercycle covariance.

In terms of the actual difference in the magnitude of interindividual variation, sampling of just two rather than six cycles per woman entails about 10% rise in variation, resulting in a 10% reduction of effect size and a consequent rise in the probability of Type II error (Cohen, 1988). Sampling three versus six cycles leads to a 5% lower effect size. Importantly, we may assume that the sample size requirements for nonseasonal populations are easier to fulfill and even a single cycle per woman would be sufficient to provide adequate statistical power. However, this liberal conclusion is contingent upon the use of full hormonal profiles, rather than single-day snapshots or similarly insufficient samples.

Number of days per cycle for the assessment of the intracycle variation

The intracycle (among-days of a single cycle) variability in progesterone levels constituted the largest source of variation in our data: the residual component accounted for 64.5% of total variance, while the two levels in the nested design ("interindividual" and "intercycle within women") together accounted for only 35.5% of total variance (Table 2). Among the 114 computed measures of within-cycle amplitude (pooling all measured cycles from 22 women), values ranged from 2.2 to 22.7 times ($n = 114$, mean 6.37, median 4.90), which means that during some cycles progesterone concentrations varied almost 23-fold (Fig. 4, filled bars). Using conservatively only days −12 to −3 of the luteal phase (thus excluding potentially very low hormone levels), the within-cycle amplitudes ranged from 1.7 to 13.8 (mean 4.18, median 3.58) (Fig. 4, open bars). In the group of 185 urban and rural women, the within-cycle amplitudes of progesterone concentrations (for days −14 to −1) ranged from 2.2 to 49.3 times (mean 8.29, median 6.20). These results underscore the importance of sufficient sampling of progesterone concentrations during individual cycles, especially during the luteal phase of the cycle.

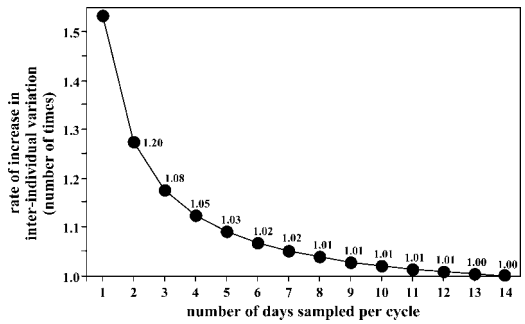

Fig. 5. The increase in interindividual variation as a function of the number of days sampled per each cycle. The magnitude of increase was scaled with respect to the level of variation for the case in which all 14 days in the cycle have been measured (sampled). The number next to each symbol shows how quickly interindividual variation decreases with each additional day measured per cycle, e.g. measuring 2 rather than 1 day per cycle reduces interindividual variation 1.20 times, and measuring 3 rather than 2 days per cycle reduces interindividual variation an additional 1.08 times.

Resampling analyses performed on data from the rural and urban populations suggest that the requirements for a reliable assessment of progesterone production are quite stringent: at least 7 or 8 days per 14-day luteal phase of the cycle should be measured (Fig. 5). The sampling of fewer than 5 days per cycle increases the level of variation

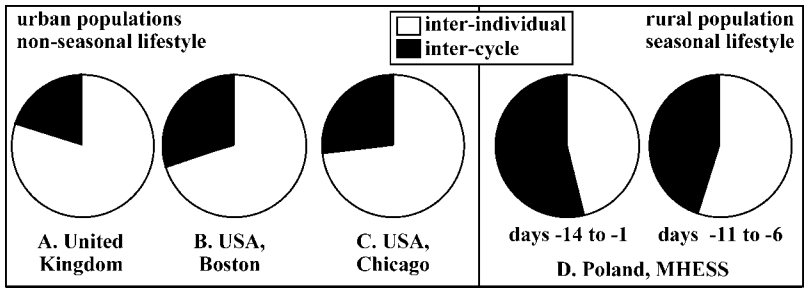

Fig. 6. Variance components associated with interindividual and intercycle (within-women) sources of variation in progesterone levels. **A.** Data published by Lenton et al. (1983) on 17 women (2 cycles per woman) from the United Kingdom. **B.** Data published by Sukalich et al. (1994) on 12 women (4 cycles per woman) from the Boston area, USA. **C.** Data published by Gann et al. (2001) on 12 women (2 cycles per woman) from the Chicago area, USA. **D.** Results of this study (22 women, on average 5.2 cycles [from 3 to 6] per woman) conducted at the Mogielica Human Ecology Study Site [MHESS], Poland.

in the data by more than 10%, thus reducing the effect size by this fraction and substantially lowering statistical power.

DISCUSSION

Determinants of hormonal variation

Levels of sex steroid hormones in premenopausal women are influenced by many factors: genes, early developmental conditions and adult lifestyle. Influence of adult lifestyle has been most intensely studied and is best understood. In particular, changes in energetic condition of an individual caused by low calorie diet, weight loss, or increased energy expenditure were related to reproductive suppression (reduced levels of hormones, inadequate luteal phase, anovulatory cycles, oligomenorrhea, and amenorrhea) in premenopausal women (Bullen et al., 1985; Chen and Brzyski, 1999; De Souza, 2003; De Souza et al., 1998; Ellison and Lager, 1986; Jasienska and Ellison, 1998, 2004; Jasienska et al., 2006c; Lager and Ellison, 1990; Morris et al., 1999; Panter-Brick and Ellison, 1994; Rosetta et al., 1998; Warren and Perlroth, 2001; Williams et al., 1999).

In women from the United States, a loss of as little as 2 kg of body weight correlated with reduced salivary progesterone levels, even though after the weight loss, women were still in comparably good nutritional conditions (Lager and Ellison, 1990). Participation in recreational sport also resulted in suppressed levels of ovarian steroids. For example, college-age women had reduced levels of salivary progesterone when jogging on average for 3 h a week (Ellison and Lager, 1986). Furthermore, in Polish urban and rural women, reduced levels of estradiol in menstrual cycles were related to high levels of daily, habitual activity (Jasienska et al., 2006c).

Variation among women in hormonal levels may also result from differences in developmental conditions (Ellison, 1996). We recently documented that size at birth, which is an indicator of energetic conditions during intrauterine development positively correlates with levels of estradiol in menstrual cycles (Jasienska et al., 2006b). Menarcheal age, which to some extent reflects conditions during childhood growth and development, shows a relationship with levels of steroid hormones in menstrual cycles (Apter, 1996; Vihko and Apter, 1984). In addition, polymorphism in genes involved in steroid metabolism has been linked to variation in levels of ovarian hormones (Feigelson et al., 1998; Jasienska et al., 2006a; Sharp et al., 2004; Small et al., 2005). Clinical relevance of intra-woman (among-cycles) variation in mean steroid hormone levels has been documented in a study showing that conception cycles were characterized by higher levels of estradiol than nonconception cycles of the same woman (Lipson and Ellison, 1996).

Seasonality and intercycle variation

Energetic factors, especially for women whose lifestyle is characterized by seasonal changes, are most likely to account for significant proportion of variation in ovarian function (Ellison, 1994; Jasienska and Thune, 2001). In rural Congo (Ellison et al., 1986) and Nepal (Panter-Brick and Ellison, 1994), women had suppressed levels of ovarian steroid hormones during seasons when they lost weight due to low caloric intake or high energy expenditure. In rural Poland, increase in energy expenditure imposed by requirements of harvest season resulted in suppressed levels of progesterone, even though increase in workload was not associated with weight loss or reduction in body fat (Jasienska and Ellison, 1998, 2004). However, even during periods when women from Nepal, Congo, Bolivia, and rural Poland have good nutritional status, they still have lower levels of ovarian steroids than women from the United States (Ellison et al., 1993).

Variation in hormonal production may therefore be expected among women from one population, among different menstrual cycles of the same woman, and finally, among populations. In women from the United States (Fig. 6B,C), intercycle variance in salivary progesterone levels accounted for about only 30% of total variance in luteal phase progesterone (Sukalich et al., 1994) and 27% of variance of peak progesterone and cumulative progesterone from 8 days of the luteal phase (Gann et al., 2001). Intercycle variation accounted for about 20% of total variation in serum progesterone levels (Fig. 6A) assessed for seven consecutive cycle-days in British women (Lenton et al., 1983). In contrast, our estimate of intercycle variance is substantially higher and accounts for 54% of variation (Fig. 6D).

Such high intercycle variation may be explained by previously published findings showing that lifestyle of Polish women, who were subjects of this study, was characterized by substantial seasonal variation in intensity of physical work, a factor known to impact reproductive physiology (Jasienska and Ellison, 1998, 2004). It is therefore likely that women experience seasonal changes in levels of reproductive hormones paralleling changes in lifestyle conditions (Ellison et al., 1989; Jasienska and Ellison,

1998, 2004; Panter-Brick and Ellison, 1994). Although data about relevant lifestyle factors (Ellison, 2003a; Jasienska and Ellison, 1998) were not provided in other studies of interindividual and intercycle variation (Gann et al., 2001; Lenton et al., 1983; Sukalich et al., 1994), it can be expected that higher intercycle variance estimated for the Polish rural women resulted from more seasonally variable lifestyle conditions than those characteristic of the urban setting.

Comparison between results of our study and those of studies on urban women (Gann et al., 2001; Lenton et al., 1983; Sukalich et al., 1994) provides preliminary support to our hypothesis that the pattern of partitioning of variation in levels of ovarian steroids is different in urban versus rural populations. In all three studies on women from urban population, the intercycle variation was low and accounted for at most 30% of variation. In our study of a rural population, the component of intercycle variation was almost two times higher. However, even in urban settings where seasonality is less likely to play a role as a determinant of lifestyle changes, attention should be paid to dieting and exercise as modern lifestyle factors known to affect ovarian physiology. For example, hormonal levels are expected to be reduced in a woman who recently lost weight, in comparison to her other cycles during which her body weight remained stable (Lager and Ellison, 1990). Therefore, a woman who usually produces high levels of hormones may be misclassified as having ovarian disturbances (e.g., insufficient luteal phase or anovulatory cycle), just because her cycle was sampled at the time when weight loss occurred.

Methodological issues: how many measurements?

It is worth emphasizing that knowledge of lifestyle conditions of a population is crucial before a decision could be made if one or more cycles measured per woman provides a reliable estimate of her hormonal status. Our exploratory analyses of variance based on data from a population characterized by pronounced seasonality yielded an extremely broad (20-fold) spectrum of ICC values (variance among individual women), which renders questionable results of studies relying on few cycles per woman. It can, however, be expected that in urban populations, such as those studied by Lenton et al. (1983), Sukalich et al. (1994), and Gann et al. (2001), characterized by less pronounced seasonal changes in lifestyle conditions, the intercycle variation would be lower than that described in our study. In sedentary, stable-weight women, their hormonal levels should not change substantially from cycle to cycle, thus lowering the required number of measured cycles.

The consequences of using insufficient sample sizes are clear (see e.g., Jasienski, 1996): elevated level of noise (measured by variance or CV) directly translates into a reduction in effect size, which is an important determinant of statistical power (Cohen, 1988). To keep the desired levels of power requires either an increase in sample size, a change in the structure of the data (such as sampling more cycles per woman), or a change in the nature of the statistical test (e.g., in multiple regression analysis, when it may result in the necessity of removing some of the independent variables from the model).

Insufficient sampling at the intracycle level may in principle be compensated by an increase in the number of subjects or the number of cycles per woman, a frequent research strategy in epidemiology. A sample of just one day per cycle reduces the effect size by as much as 50–60% (compared to a full 14-measurement luteal phase); maintaining the same power of the t test comparing two groups of women can only be compensated by a fourfold increase in the number of subjects (Cohen, 1988). In some instances, this trade-off may be justified by substantially lower research costs or logistical convenience.

However, lack of knowledge about the intracycle dynamics of progesterone production may render more refined analyses impossible. Similarly, replacing it with simplistic and potentially misleading ratios of two extreme values of hormone concentrations may make the ratio-based models statistically untestable (Jasienski and Bazzaz, 1999). Modeling the hormonal profiles during the menstrual cycles as function-valued traits (e.g. Kirkpatrick and Meyer, 2004) or a search for powerful smoothing models (Brumback and Rice, 1998) may yield novel parameters of explanatory potential, but both cases require full information about daily hormone levels. Of more immediate concern, however, is the need to study the entire luteal phase to properly evaluate the timing of the mid-luteal phase. For example, there are reasons to think that the mid-luteal events may carry more biologically meaningful information than data averaged across the entire menstrual cycle (Jasienska, unpublished). The "interindividual" component is particularly pronounced for days -11 through -6, which correspond to the mid-luteal phase. Including the entire luteal phase (days -14 to -1) in analyses introduces a component of random variation among cycles, thus potentially elevating the intercycle component (and reducing the interindividual component). This finding may have important implications for reproductive ecology and epidemiology since mid-luteal hormonal levels are likely to exhibit more meaningful or robust correlations with life-historical variables.

The knowledge of the levels of indigenous ovarian steroids of a woman is of unquestionable importance in clinical practice and public health. Individual assessment of hormone levels during menstrual cycle could be important in using these values as biomarkers of risk of hormone-depended cancers. Lifetime levels of estrogens and progesterone are hypothesized to play a crucial role in the development of breast and reproductive cancers in women (Bernstein and Ross, 1993; Jasienska et al., 2000; Pike et al., 1993). If they are to serve as reliable biomarkers, however, it is essential to take into account the existence of genuine, nonpathological, variation in hormone levels, both among individual women and among cycles of a healthy individual woman.

ACKNOWLEDGMENTS

GJ acknowledges generous support from the Radcliffe Fellowship 2005–2006 program. GJ and MJ are grateful to the anonymous referee for helpful insights. This is an equal authorship paper.

LITERATURE CITED

Apter D. 1996. Hormonal events during female puberty in relation to breast cancer risk. Eur J Cancer Prev 5:476–482.

Bernstein L, Ross RK. 1993. Endogenous hormones and breast cancer risk. Epidemiol Rev 15:48–65.

Brumback BA, Rice JA. 1998. Smoothing spline models for the analysis of nested and crossed samples of curves. J Am Stat Assoc 93:961–976.

Bullen BA, Skrinar GS, Beitins IZ, von Mering G, Turnbull BA, McArthur JW. 1985. Induction of menstrual disorders by strenuous exercise in untrained women. N Engl J Med 312:1349–1353.
Chen EC, Brzyski RG. 1999. Exercise and reproductive dysfunction. Fertil Steril 71:1–6.
Cohen J. 1988. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: Erlbaum.
De Souza MJ. 2003. Menstrual disturbances in athletes: A focus on luteal phase defects. Med Sci Sports Exerc 35:1553–1563.
De Souza MJ, Miller BE, Loucks AB, Luciano AA, Pescatello LS, Campbell CG, Lasley BL. 1998. High frequency of luteal phase deficiency and anovulation in recreational women runners: blunted elevation in follicle-stimulating hormone observed during luteal-follicular transition. J Clin Endocr Metab 83:4220–4232.
DeBruine LM, Jones BC, Perrett DI. 2005. Women's attractiveness judgments of self-resembling faces change across the menstrual cycle. Horm Behav 47:379–383.
Ellison PT. 1994. Salivary steroids and natural variation in human ovarian function. Ann NY Acad Sci 709:287–298.
Ellison PT. 1996. Developmental influences on adult ovarian hormonal function. Am J Hum Biol 8:725–734.
Ellison PT, editor. 2001. Reproductive ecology and human evolution. Chicago: Aldine De Gruyter.
Ellison PT. 2003a. Energetics and reproductive effort. Am J Hum Biol 15:342–351.
Ellison PT. 2003b. On fertile ground. Cambridge, MA: Harvard University Press.
Ellison PT, Lager C. 1986. Moderate recreational running is associated with lowered salivary progesterone profiles in women. Am J Obstet Gynecol 154:1000–1003.
Ellison PT, Peacock NR, Lager C. 1986. Salivary progesterone and luteal function in two low-fertility populations of northeast Zaire. Hum Biol 58:473–483.
Ellison PT, Peacock NR, Lager C. 1989. Ecology and ovarian function among Lese women of Ituri Forest Zaire. Am J Phys Anthropol 78:519–526.
Ellison PT, Panter Brick C, Lipson SF, O'Rourke MT. 1993. The ecological context of human ovarian function. Hum Reprod 8:2248–2258.
Feigelson HS, Shames LS, Pike MC, Coetzee GA, Stanczyk FZ, Henderson BE. 1998. Cytochrome p450c17 alpha gene (CYP17) polymorphism is associated with serum estrogen and progesterone concentrations. Cancer Res 58: 585–587.
Feinberg DR, Jones BC, Law-Smith MJ, Moore FR, DeBruine LM, Cornwell RE, Hillier SG, Perrett DI. 2006. Menstrual cycle trait estrogen level and masculinity preferences in the human voice. Horm Behav 49:215–222.
Ferrell RJ, O'Connor KA, Rodriguez G, Gorrindo T, Holman DJ, Brindle E, Miller RC, Schechter DE, Korshalla L, Simon JA, Mansfield PK, Wood JW, Weinstein M. 2005. Monitoring reproductive aging in a 5-year prospective study: Aggregate and individual changes in steroid hormones and menstrual cycle lengths with age. Menopause 12:567–577.
Gann PH, Giovanazzi S, Van Horn L, Branning A, Chatterton RT. 2001. Saliva as a medium for investigating intra- and interindividual differences in sex hormone levels in premenopausal women. Cancer Epidemiol Biomark Prev 10:59–64.
Hillard PJA, Nelson LM. 2003. Adolescent girls the menstrual cycle and bone health. J Pediatr Endocr Metab 16:673–681.
Himmelstein DU, Levins R, Woolhandler S. 1990. Beyond our means: patterns of variability of physiological traits. Int J Health Serv 20:115–124.
Holman DJ, Grimes MA, Brindle E, O'Connor KA. 2004. Hormonal correlates for the initiation of breast-feeding in Bangladeshi women. Horm Behav 46:382–391.
Jasienska G. 2003. Energy metabolism and the evolution of reproductive suppression in the human female. Acta Biotheor 51:1–18.
Jasienska G, Ellison PT. 1998. Physical work causes suppression of ovarian function in women. Proc R Soc Lond B Biol Sci 265:1847–1851.
Jasienska G, Ellison PT. 2004. Energetic factors and seasonal changes in ovarian function in women from rural Poland. Am J Hum Biol 16:563–580.
Jasienska G, Kapiszewska M, Ellison PT, Kalemba-Drozdz M, Nenko I, Thune I, Ziomkiewicz A. 2006a. CYP17 genotypes differ in salivary 17-β estradiol levels: A study based on hormonal profiles from entire menstrual cycles. Cancer Epidemiol Biomark Prev 15:2131–2135.
Jasienska G, Thune I. 2001. Lifestyle hormones and risk of breast cancer. Br Med J 322:586–587.
Jasienska G, Thune I, Ellison PT. 2000. Energetic factors ovarian steroids and the risk of breast cancer. Eur J Cancer Prev 9:231–239.
Jasienska G, Ziomkiewicz A, Ellison PT, Lipson SF, Thune I. 2004. Large breasts and narrow waists indicate high reproductive potential in women. Proc R Soc Lond B Biol Sci 271:1213–1217.
Jasienska G, Ziomkiewcz A, Lipson SF, Thune I, Ellison PT. 2006b. High ponderal index at birth predicts high estradiol levels in adult women. Am J Hum Biol 18:133–140.
Jasienska G, Ziomkiewicz A, Thune I, Lipson SF, Ellison PT. 2006c. Habitual physical activity and estradiol levels in women of reproductive age. Eur J Cancer Prev 15:439–445.
Jasienski M. 1996. N=1 fortuitous effects and AIDS research. The Scientist 10:12.
Jasienski M, Bazzaz FA. 1999. The fallacy of ratios and the testability of models in biology. Oikos 84:321–325.
Jones BC, Little AC, Boothroyd L, DeBruine LM, Feinberg DR, Law Smith MJ, Cornwell RE, Moore FR, Perrett DI. 2005. Commitment to relationships and preferences for femininity and apparent health in faces are strongest on days of the menstrual cycle when progesterone level is high. Horm Behav 48:283–290.
Kaaks R, Berrino F, Key T, Rinaldi S, Dossus L, Biessy C, Secreto G, Amiano P, Bingham S, Boeing H, de Mesquita HBB, Chang-Claude J, Clavel-Chapelon FO, Fournier AS, van Gils C, Gonzalez CA, Gurrea AB, Critselis E, Khaw KT, Krogh V, Lahmann PH, Nagel G, Olsen A, Onland-Moret NC, Overvad K, Palli D, Panico S, Peeters P, Quiros JR, Roddam A, Thiebaut A, Tjonneland A, Chirlaque MD, Trichopoulou A, Trichopoulos D, Tumino R, Vineis P, Norat T, Ferrari P, Slimani N, Riboli E. 2005. Serum sex steroids in premenopausal women and breast cancer risk within the European prospective investigation into cancer and nutrition (EPIC). J Natl Cancer Inst 97:755–765.
Kirkpatrick M, Meyer K. 2004. Direct estimation of genetic principal components: simplified analysis of complex phenotypes. Genetics 168:2295–2306.
Lager C, Ellison PT. 1990. Effect of moderate weight loss on ovarian function assessed by salivary progesterone measurements. Am J Hum Biol 2:303–312.
Lenton E, Lawrence G, Coleman R, Cooke I. 1983. Individual variation in gonadotrophin and steroid concetration and in lengths of the follicular and luteal phases in women with regular menstrual cycles. Clin Reprod Fertil 2:143–150.
Lipson SF, Ellison PT. 1992. Normative study of age variation in salivary progesterone profiles. J Biosoc Sci 24: 233–244.
Lipson SF, Ellison PT. 1996. Comparison of salivary steroid profiles in naturally occurring conception and non-conception cycles. Hum Reprod 11:2090–2096.
Mead LA, Hampson E. 1997. Turning bias in humans is influenced by phase of the menstrual cycle. Horm Behav 31:65–74.
Morris FL, Payne WR, Wark JD. 1999. Prospective decrease in progesterone concentrations in female lightweight lowers during the competition season compared with the off season: a controlled study examining weight loss and intensive exercise. Br J Sports Med 33:417–422.
Noh JJ, Maskarinec G, Pagano I, Cheung LWK, Stanczyk FZ. 2006. Mammographic densities and circulating hormones: a cross-sectional study in premenopausal women. Breast 15:20–28.
Panter-Brick C, Ellison PT. 1994. Seasonality of workloads and ovarian function in Nepali women. Ann NY Acad Sci 709:234–235.
Pawlowski B, Jasienska G. 2005. Women's preferences for sexual dimorphism in height depend on menstrual cycle phase and expected duration of relationship. Biol Psychol 70:38–43.
Pike MC, Spicer DV, Dahmoush L, Press MF. 1993. Estrogens progestogens normal breast cell proliferation and breast cancer risk. Epidemiol Rev 15:17–35.
Rosetta L, Harrison GA, Read GF. 1998. Ovarian impairments of female recreational distance runners during a season of training. Ann Hum Biol 25:345–357.
Schultheiss OC, Dargel A, Rohde W. 2003. Implicit motives and gonadal steroid hormones: effects of menstrual cycle phase oral contraceptive use and relationship status. Horm Behav 43:293–301.
Sharp L, Cardy AH, Cotton SC, Little J. 2004. CYP17 gene polymorphisms: prevalence and associations with hormone levels and related factors. A HuGE review. Am J Epidemiol 160:729–740.
Sherwin BB. 2005. Estrogen and memory in women: How can we reconcile the findings? Horm Behav 47:371–375.
Small CM, Marcus M, Sherman SL, Sullivan AK, Manatunga AK, Feigelson HS. 2005. CYP17 genotype predicts serum hormone levels among pre-menopausal women. Hum Reprod 20:2162–2167.
Sukalich S, Lipson S, Ellison PT. 1994. Intra and interwomen variation in progesterone profiles. Am J Phys Anthropol 18(Suppl):191.
Vihko R, Apter D. 1984. Endocrine characteristics of adolescent menstrual cycles—Impact of early menarche. J Steroid Biochem Mol Biol 20:231–236.
Vitzthum VJ, Ellison PT, Sukalich S. 1994. Salivary progesterone profiles of indigenous Andean women. Am J Phys Anthropol 18(Suppl):201–202.
Warren MP, Perlroth NE. 2001. The effects of intense exercise on the female reproductive system. J Endocrinol 170:3–11.
Williams CL. 1998. Estrogen effects on cognition across the lifespan. Horm Behav 34:80–84.
Williams NI, Bullen BA, McArthur JW, Skrinar GS, Turnbull BA. 1999. Effects of short-term strenuous endurance exercise upon corpus luteum function. Med Sci Sports Exerc 31:949–958.