Serum caffeine and paraxanthine concentrations and menstrual cycle function: correlations with beverage intakes and associations with race, reproductive hormones, and anovulation in the BioCycle Study[1,2]

*Karen C Schliep,[3,4] Enrique F Schisterman,[3] Jean Wactawski-Wende,[5] Neil J Perkins,[3] Rose G Radin,[3] Shvetha M Zarek,[3,6] Emily M Mitchell,[3] Lindsey A Sjaarda,[3] and Sunni L Mumford[3]**

[3]Epidemiology Branch, Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Rockville, MD; [4]Department of Family and Preventive Medicine, University of Utah Health Sciences Center, Salt Lake City, UT; [5]Department of Epidemiology and Environmental Health, University at Buffalo, Buffalo, NY; and [6]Program of Reproductive and Adult Endocrinology, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Bethesda, MD

ABSTRACT

Background: Clinicians often recommend limiting caffeine intake while attempting to conceive; however, few studies have evaluated the associations between caffeine exposure and menstrual cycle function, and we are aware of no previous studies assessing biological dose via well-timed serum measurements.

Objectives: We assessed the relation between caffeine and its metabolites and reproductive hormones in a healthy premenopausal cohort and evaluated potential effect modification by race.

Design: Participants ($n = 259$) were followed for ≤2 menstrual cycles and provided fasting blood specimens ≤8 times/cycle. Linear mixed models were used to estimate associations between serum caffeine biomarkers and geometric mean reproductive hormones, whereas Poisson regression was used to assess risk of sporadic anovulation.

Results: The highest compared with the lowest serum caffeine tertile was associated with lower total testosterone [27.9 ng/dL (95% CI: 26.7, 29.0 ng/dL) compared with 29.1 ng/dL (95% CI: 27.9, 30.3 ng/dL), respectively] and free testosterone [0.178 ng/mL (95% CI: 0.171, 0.185 ng/dL) compared with 0.186 ng/mL (95% CI: 0.179, 0.194 ng/dL), respectively] after adjustment for age, race, percentage of body fat, daily vigorous exercise, perceived stress, depression, dietary factors, and alcohol intake. The highest tertiles compared with the lowest tertiles of caffeine and paraxanthine were also associated with reduced risk of anovulation [adjusted RRs (aRRs): 0.39 (95% CI: 0.18, 0.87) and 0.40 (95% CI: 0.18, 0.87), respectively]. Additional adjustment for self-reported coffee intake did not alter the reproductive hormone findings and only slightly attenuated the results for serum caffeine and paraxanthine and anovulation. Although reductions in the concentrations of total testosterone and free testosterone and decreased risk of anovulation were greatest in Asian women, there was no indication of effect modification by race.

Conclusion: Caffeine intake, irrespective of the beverage source, may be associated with reduced testosterone and improved menstrual cycle function in healthy premenopausal women. *Am J Clin Nutr* 2016;104:155–63.

Keywords: anovulation, biological markers, caffeine, menstrual cycle, 1,7-dimethylxanthine

INTRODUCTION

Caffeine is consumed daily by the majority of US women and is ingested via natural and supplemented sources (1, 2). Despite the common counsel for women to either reduce or entirely eliminate caffeine intake while attempting to conceive (3), few studies have evaluated the association between caffeine exposure and menstrual cycle function and fertility, and available studies have produced inconsistent results (4). Inferences of findings from these studies have been hampered by methodologic limitations such as an inadequate hormone assessment with the use of standard methods to time a woman's cycle phase, inadequate control for confounding variables, imprecise caffeine-exposure assessment, and failure to account for individual and cycle phase differences in caffeine clearance (4).

The ideal measure of female exposure to caffeine and its metabolites, when their effects on menstrual cycle function are assessed, is via serum samples during relevant times including the follicular, ovulatory, and luteal phases (5). However, studies to date that have assessed caffeine's effects on menstrual cycle function in nonpregnant women have relied on self-reported intake (6–10). Reported dietary assessments, particularly if collected retrospectively with respect to an adverse outcome such as an inability to achieve a pregnancy, are subject to recall bias (5). Furthermore, even a perfect report of caffeinated beverage,

[1] Supported by the Intramural Research Program, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, NIH (contracts HHSN275200403394C and HHSN275201100002I and Task 1 HHSN27500001) and the State University of New York at Buffalo. The NIH Office of Dietary Supplements (OD-00-113) provided support for the caffeine-biomarker assessment.

[2] Supplemental Figures 1 and 2 and Supplemental Tables 1–6 are available from the "Online Supporting Material" link in the online posting of the article and from the same link in the online table of contents at http://ajcn.nutrition.org.

*To whom correspondence should be addressed. E-mail: mumfords@mail.nih.gov.

Received July 1, 2015. Accepted for publication April 25, 2016.

First published online May 25, 2016; doi: 10.3945/ajcn.115.118430.

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018

food, and medication intakes may have substantial measurement error in the biological dose. This measurement error is due to the variability in the caffeine contents of coffee and tea, which are known to differ significantly by roasting and brewing methods even within the same barista (2); and the intervariability and intravariability in caffeine metabolism, which is known to be affected by genetic and nongenetic factors such as smoking, oral contraceptives, race, and menstrual cycle phase (4, 11). Because of the potential cost and logistic considerations for measuring caffeine biomarkers, researchers are interested in identifying whether self-report of caffeine intake adequately agrees with the biomarker gold standard (12) particularly in premenopausal women in whom intrawoman caffeine metabolism may vary over the menstrual cycle (11, 13, 14).

We previously showed that, in healthy premenopausal women, self-reported caffeine intake was inversely associated with free estradiol in whites and positively associated with free estradiol in Asians, but that there were no associations between caffeine and ovulatory function in all races (15). In this project, with the use of data from the same study population but augmented with fasting morning serum caffeine biomarkers and testosterone, we set out to fill previous data gaps by determining whether serum caffeine and its chief metabolites paraxanthine and theobromine are associated with reproductive hormones and ovulatory function while also evaluating a potential effect modification by race. To inform future research in women of reproductive age, we further assessed whether reported caffeine intake adequately agrees with serum caffeine biomarkers while taking into account potential differences in caffeine metabolism such as race, smoking status, and menstrual cycle phase.

METHODS

Study population

The BioCycle Study, which was conducted in 2005–2007, followed 259 women from New York for ≤2 menstrual cycles. The study population, materials, and methods have been previously described (16). In summary, the convenience sample consisted of carefully screened healthy women between the ages of 18 and 44 y who had no history of chronic diseases such as heart disease, diabetes mellitus, cancer, inflammatory diseases, autoimmune, liver, or kidney disease, thyroid disease, or any other endocrine dysfunction. Furthermore, included women had to be regularly menstruating (self-reported cycle length between 21 and 35 d for each menstrual cycle in the past 6 mo) to participate. Women with conditions known to affect menstrual cycle function such as polycystic ovary disease, uterine fibroids, or current use of hormonal contraception (i.e., 3 mo before study entry) were excluded as were women who were currently underweight or overweight or planning to restrict their diet, had a recent history of alcohol abuse or illicit drug use, or were consuming a high soy–based diet. The Health Sciences Institutional Review Board at the University at Buffalo approved the study and served as the review board designated by the NIH under a reliance agreement.

Serum caffeine, paraxanthine, and theobromine assessments

Paraxanthine is the primary metabolite of caffeine (84%) (17) with the lesser metabolites being theobromine (12%) and theophylline (4%). The latter 2 metabolites are also present in their whole form in chocolate and tea leaves and, compared with paraxanthine, have been shown to be weakly correlated with caffeine. For these analyses, we assessed serum caffeine and the 2 major metabolites paraxanthine and theobromine. A modification of a method that was used to analyze caffeine and metabolites in serum and tissues via the 1100 liquid chromatography–mass spectrometry detector (Agilent) was used for the analysis of caffeine, paraxanthine, and theobromine. The chromatographic conditions permitted an excellent separation of paraxanthine from 2 other commonly encountered caffeine metabolites (theophylline and theobromine), which have the same mass. To reduce measurement error, select batches of measurements were recalibrated postassay by a calibration curve estimated from all calibration data (18). With the use of recalibrated data, the interassay CVs were <10% for caffeine (an improvement of 10% compared with laboratory values) and <19% for theobromine (an improvement of 15% compared with laboratory values). For paraxanthine, there was no obvious advantage that would be gained by recalibration (overall or in particular batches), and thus, original laboratory values were used.

Self-reported caffeine assessment

Participants underwent a 24-h dietary recall (24HDR)[6] at the clinic after the collection of fasting blood specimens during 4 cycle visits corresponding to menstruation, the midfollicular phase, ovulation, and the midluteal phase (15). Cycle visits were scheduled between 0700 and 0830 to allow for the collection of fasting samples and to reduce the diurnal variation. Food and beverage intakes for the previous day, with estimated times of intake recorded, were assessed with the use of the standardized, multiple-pass approach. Nutrient data were analyzed with the use of the Nutrition Data System for Research (version 2005; Nutrition Coordinating Center, University of Minnesota) with an estimated caffeine content per 8 oz (237 mL) of coffee equaling 96 mg. Eighty-seven percent of participants completed four 24HDRs/cycle.

Reproductive hormone and sporadic anovulation assessments

Eligible women who consented to participate provided fasting blood specimens at ≤8 visits/cycle for 2 menstrual cycles with visits timed with the use of fertility monitors to correspond with menstruation, midfollicular phase, late follicular phase, luteinizing hormone (LH)/follicle-stimulating hormone (FSH) surge, ovulation, and early luteal, midluteal, and late-luteal phases. Collection dates were adjusted for cycle length with the use of Clearblue fertility monitors (SPD Swiss Precision Diagnostics), which are useful in the timing of clinic visits and have been shown to improve cycle standardization. Fertility monitors aid in biologically standardizing cycles that vary considerably from the 28-d standard cycle (19). Thus, the use of the fertility monitors helped to align the participant cycles so that, despite their cycle lengths, hormone concentrations could be compared across women at

[6] Abbreviations used: aRR, adjusted RR; FSH, follicle-stimulating hormone; LH, luteinizing hormone; 24HDR, 24-h dietary recall.

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018

various cycle phases. A total of 94% of participants completed 7–8 clinic visits/cycle.

Total estradiol, progesterone, LH, FSH, and sex hormone binding globulin were measured with the use of a solid-phase, competitive, chemiluminescent enzymatic immunoassay on a DPC IMMULITE 2000 analyzer (Siemens Medical Solutions Diagnostics) at Kaleida Laboratories (Buffalo, New York). Total testosterone was measured via liquid chromatography–tandem mass spectrometry with the use of a Shimadzu Prominence Liquid Chromatogram (Shimadzu Scientific Instruments Inc.) with an ABSceix 5500 tandem mass spectrometer (AB SCIEX) at the Advanced Research and Diagnostics Laboratory (Minneapolis, Minnesota). The calculation of free (i.e., bioavailable) estradiol and testosterone was performed via standardized methods (20, 21). Across the study period, CVs for these tests were as follows: <14% for progesterone; <10% for estradiol, testosterone, and sex hormone binding globulin; and <5% for LH and FSH. We defined anovulation as any cycle with a peak progesterone concentration <5 ng/mL and no observed serum LH peak at the midluteal or late-luteal phase visits (n = 42 of 509 cycles; 8.3%) (22).

Covariate assessment

Age, race, smoking status, and reproductive history were obtained at baseline with the use of standard questionnaires (16). Depression was also assessed at baseline with the use of the 20-item Center for Epidemiologic Studies Depression scale. Participants were asked to measure their frequencies of 20 depressive feelings and behaviors in the past week with responses ranging from "rarely or none of the time" (0 points) to "most or all the time" (3 points). Each participant was given an overall score (range: 0–60) (23). At the end of the follow-up period, the total percentage of body fat was measured with the use of dual-energy X-ray absorptiometry (Hologic Discovery Elite, software version 12.4.1; Hologic Inc.). Total energy (kcal/d), fiber (g/d), and coffee intake (cups/d) and the alternate Mediterranean diet score were calculated from the 24HDR (24).

Daily minutes of vigorous exercise, alcohol intake (total number of drinks consumed over the course of the day, noting 0 if none), perceived stress [not stressful (1), a little stressful (2), or very stressful (3)], and any pain medication use over the course of the day (including aspirin, naproxen, ibuprofen, and acetaminophen) were captured via a daily diary. Caffeine from medications (primarily from over-the-counter preparations with nonsteroidal anti-inflammatory drugs) was averaged for each phase over the 2 cycles and added to the caffeine that was calculated from the eight 24HDRs (15).

Statistical analyses

Caffeine intake, as reported in the 24HDR, and serum caffeine, paraxanthine, and theobromine were assessed both categorically (tertiles) and continuously (as mg/d for 24HDR and μmol $\cdot$ L^{-1} $\cdot$ mL^{-1} for serum caffeine and metabolites). Descriptive statistics were compared between study-mean serum caffeine tertiles and population characteristics with the use of an ANOVA for normally distributed continuous variables and the Wilcoxon-Mann-Whitney test for nonnormally distributed continuous variables. Chi-square or Fisher's exact tests, when appropriate, were used for categorical variables. Variations in serum caffeine, paraxanthine, and theobromine concentrations across the menstrual cycle were assessed with the use of linear mixed models. Pairwise comparisons between days were made with the use of Tukey's method to account for multiple comparisons.

Linear mixed models with random intercepts were used to assess the associations between serum caffeine, paraxanthine, and theobromine and reproductive hormones. Because of skewed distributions, serum caffeine and metabolites were categorized into tertiles, and reproductive hormone concentrations were log transformed for statistical analyses. These random-intercept models were chosen to account for the variation between baseline concentrations of hormones in individual women and the correlation between cycles of the same women. We calculated the P-trend value by taking the median of each serum caffeine, paraxanthine, and theobromine tertile and analyzing this as a continuous variable. Multivariable generalized linear models and Poisson regression with robust error variance (taking into account within-woman repeated cycles) were used to assess the RR of sporadic anovulation. Covariates to include in our adjusted analyses were determined by a review of the previous literature and statistical testing for confounding identification. Variables retained in the final multivariable models included age (continuous), race (white, black, Asian, or other), percentage of body fat (continuous), depression (continuous), time-varying reported alcohol and total energy intakes, Mediterranean diet score, perceived stress, and vigorous exercise (all continuous). We ran additional models with further adjustment for self-reported coffee intake (cups/d) to better understand the influence of other components in the largest contributing source of caffeine. In adjusted RR (aRR) models for sporadic anovulation, serum caffeine, paraxanthine, theobromine, and potential confounders were averaged across the cycle with caffeine biomarkers assessed via cycle-average tertiles. Because of previous findings for an effect modification by race (15), we tested for an interaction between serum caffeine biomarkers and reproductive hormones and sporadic anovulation by race and showed race-stratified results for the associations.

To determine the validity of previous day's self-reported caffeine intake (via the 24HDR) relative to fasting morning serum caffeine, paraxanthine, and theobromine, Pearson's product-moment correlation coefficients were used. Caffeine biomarkers and 24HDR reported intake were categorized dichotomously (above and below the median), and prevalence and bias-adjusted κ coefficients were calculated (25) as were the sensitivity and specificity between previous 24-h self-reported caffeine intake (test method) and serum caffeine biomarkers (the gold standard). In addition, we performed cross-classification analyses and compared the percentage of agreement and weighted κ coefficients that were calculated with a quadratic set of weights between 24HDR intake and serum biomarker quartiles. Reported caffeine intake and measured weight were recorded at baseline and were used to calculate the caffeine dosage expressed as mg $\cdot$ kg^{-1} $\cdot$ d^{-1}. These measurements were done because, at a given intake, larger individuals will have lower serum values and, consequently, a lower biological dose. Because of previous research that indicated differences in caffeine metabolism by race, smoking status, and menstrual cycle phase, an effect modification between reported intake and serum caffeine biomarkers by race (white, black, or Asian), smoking

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018

status (smoker or nonsmoker), and cycle phase (menses, early follicular phase, estimated day of ovulation, and midluteal phase) were evaluated by fitting interaction terms in multivariate models. A sensitivity analysis was performed by assessing the correlation and agreement between 24HDR self-reported caffeine intake and serum caffeine and paraxanthine by the recency of intake overall and stratified by race (i.e., <13, 13–18, >18–23, and >23–28 h). All analyses were performed in SAS version 9.4 software (SAS Institute Inc.).

RESULTS

Across the 2 menstrual cycles of the study, mean ± SD and median serum caffeine concentrations were 2.35 ± 2.74 and 1.45 μmol/L (IQR: 0.66–2.96 μmol/L; range: 0.09–21.7 μmol/L), respectively; mean ± SD and median serum paraxanthine concentrations were 1.33 ± 1.34 and 0.85 μmol/L (IQR: 0.47–1.64 μmol/L; range 0.35–12.69 μmol/L); and mean ± SD and median serum theobromine concentrations were 3.27 ± 2.84 and 2.67 μmol/L (IQR: 1.40–4.36 μmol/L; range: 0.006–23.90 μmol/L). Tertiles of serum caffeine were positively associated with age and inversely associated with the depression score. Serum caffeine was also positively associated with white race and pain-medication use (**Table 1**). Serum caffeine, paraxanthine, and theobromine showed no significant variation in pairwise comparisons across the menstrual cycle with the exception of serum caffeine, which was highest at the midfollicular visit compared with at the early follicular and early and midluteal phase visits (**Figure 1**).

Serum caffeine biomarkers and reproductive function

Assessments of serum caffeine tertiles and reproductive hormones showed significantly decreasing concentrations of total testosterone (29.1, 28.2, and 27.9 ng/mL; *P*-trend = 0.02), free testosterone (0.186, 0.180, and 0.178 ng/mL; *P*-trend = 0.02), and LH (6.89, 6.13, and 6.09 ng/mL; *P*-trend = 0.04) with increasing serum caffeine concentrations (**Table 2**) after adjustment for age, race, percentage of body fat, daily vigorous exercise, perceived stress, depression, Mediterranean diet score, and total energy and alcohol intakes. Lower total testosterone and LH were seen for middle compared with lowest serum paraxanthine tertiles [27.9 compared with 28.8 ng/mL (Tukey-adjusted $P = 0.04$) and 5.97 compared with 6.68 ng/mL (Tukey-adjusted $P = 0.008$), respectively]. Results were similar after additional adjustment for self-reported caffeinated beverage intake (**Supplemental Table 1**). No significant trend associations were shown for serum paraxanthine and theobromine and reproductive hormones nor for the testosterone:estradiol ratio. Interactions between race, smoking, and menstrual cycle phase and serum caffeine biomarkers were all nonsignificant (likelihood ratio test, all $P > 0.05$). Although there was no significant effect modification by race, the strongest associations both in magnitude and significance between increasing serum caffeine

TABLE 1
Demographic, lifestyle, and dietary characteristics of participants by mean serum caffeine intake across the study period (n = 259 women)[1]

		Serum caffeine tertile, μmol · L^{-1} · mL^{-1}			
	Overall	1 (0.11–0.89)	2 (>0.89–2.45)	3 (>2.45–18.17)	*P*
Participants, *n*	259	86	87	86	
Age, y	27.3 ± 8.2[2]	25.0 ± 6.8[a]	26.9 ± 8.6[a]	29.9 ± 8.5[b]	<0.001
Race, *n* (%)					0.005
White	154 (60)	42 (49)	50 (57)	62 (72)	
Black	51 (20)	27 (31)	15 (17)	9 (10)	
Asian	37 (14)	11 (13)	13 (15)	13 (15)	
Other	17 (7)	6 (7)	9 (10)	2 (2)	
Body fat, %	29.5 ± 5.9	28.7 ± 6.3	29.9 ± 5.4	29.9 ± 5.9	0.34
Nulliparous, *n* (%)	66 (26)	69 (82)	62 (74)	56 (66)	0.06
Depression score	5 (1–9)[3]	6 (2–9)[a]	5 (2–8)	3 (0–7)[b]	0.006
Current smoker, *n* (%)	42 (16)	9 (10)	16 (18)	17 (20)	0.20
Perceived stress	1.50 ± 0.33	1.50 ± 0.33	1.51 ± 0.33	1.48 ± 0.33	0.78
Vigorous exercise, min/d	9.2 (1.9–20.2)	5.6 (1.0–16.0)	9.8 (1.9–22.4)	10.7 (3.3–21.5)	0.11
Pain medication use, *n* (%)	175 (68)	49 (57)	63 (72)	63 (73)	0.04
Total energy, kcal/d	1613.3 ± 367.3	1656 ± 368.4	1542.2 ± 342.9	1642.5 ± 383.2	0.08
Fiber, g/d	13.6 ± 5.6	14.2 ± 6.4	12.9 ± 5.4	13.8 ± 4.9	0.29
Mediterranean diet score	2.9 ± 0.9	2.9 ± 1.0	2.7 ± 1.0	3.0 ± 0.8	0.14

[1]Descriptive statistics were compared between categories of serum caffeine averaged over the study period with the use of an ANOVA for normally distributed continuous variables and the Wilcoxon-Mann-Whitney test for nonnormally distributed continuous variables. Chi-square or Fisher's exact tests, when appropriate, were used for categorical variables. Values in a row that do not share a common superscript letter were significantly different at $P < 0.05$ via a pairwise comparison with the use of Tukey's method. There were no missing values for the following variables: age; race; percentage of body fat; smoking status; pain medication use; average alcohol, fiber, and energy intakes; vigorous exercise; Mediterranean diet score; and reproductive hormone concentrations. There were 11 missing values for the depression score and 4 missing values for parity.

[2]Mean ± SD (all such values).

[3]Median; IQR in parentheses (all such values).

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018

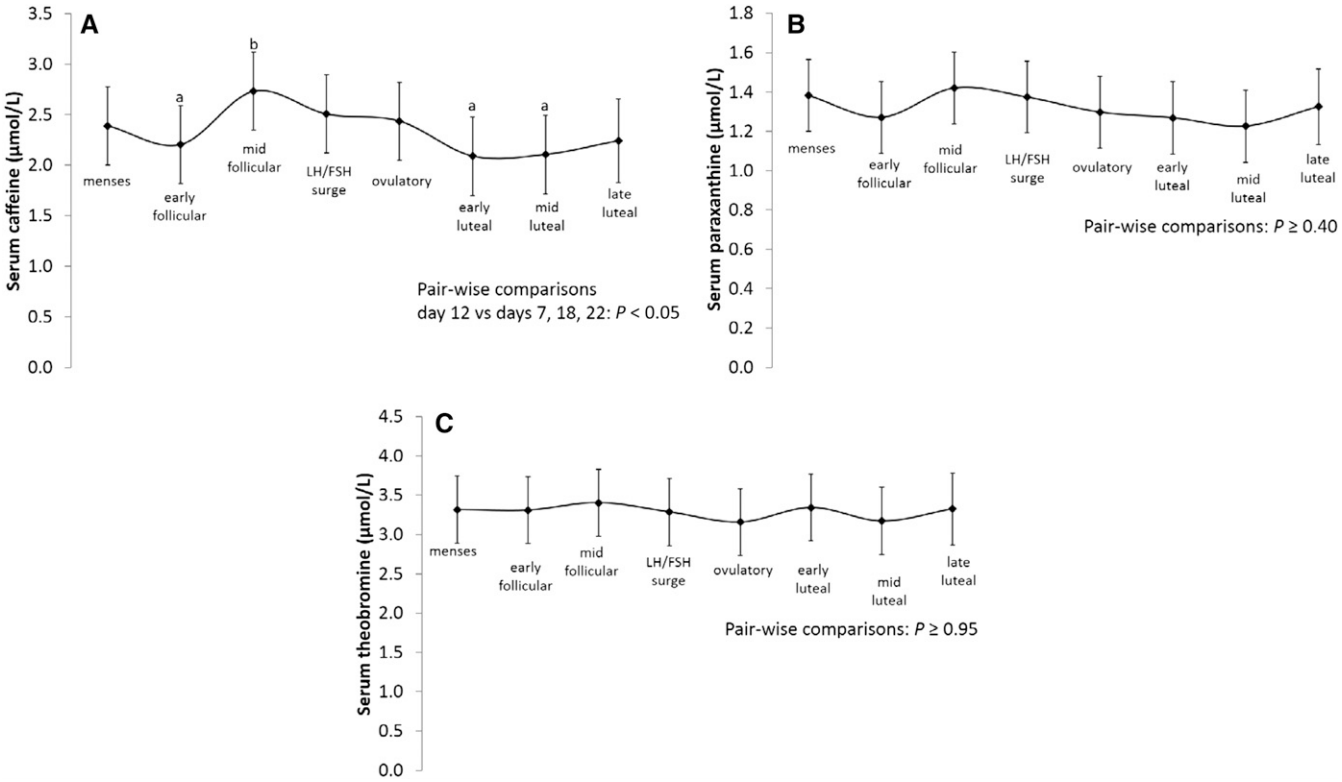

FIGURE 1 Mean ± SD variations in fasting serum caffeine (A), paraxanthine (B), and theobromine (C) concentrations across the menstrual cycle (n = 259 women). Comparisons were made with the use of linear mixed models to account for repeated measures within women (both across the cycle and between cycles). Pairwise comparisons were made between menses, early follicular, midfollicular, LH/FSH surge, ovulatory, early luteal, midluteal, and late-luteal visits with the use of Tukey's method to account for multiple comparisons. Values that do not share a common lowercase letter were significantly different at $P < 0.05$ on the basis of Tukey's test. Overall, there were ≥3650 measurements (90%) for each reproductive hormone collected from ≤16 clinic visits across 2 menstrual cycles for the 259 women. Note: 1.00 μmol serum caffeine/L = 194.2 ng/mL; 1.00 μmol serum paraxanthine/L = 180.2 ng/mL; and 1.00 μmol serum theobromine/L = 180.2 ng/mL. FSH, follicle-stimulating hormone; LH, luteinizing hormone.

and lower free testosterone, total testosterone, and LH and higher progesterone concentrations were seen in Asians (**Supplemental Figure 1**, **Supplemental Tables 2–4**).

As regards menstrual cycle function, we showed a reduced RR of anovulation for the highest tertiles of serum caffeine and paraxanthine compared with the lowest tertiles of serum caffeine and paraxanthine after adjustment for age, race, percentage of body fat, vigorous exercise, perceived stress, and total energy and alcohol intakes [aRRs: 0.39 (95% CI: 0.18, 0.87) and 0.40, (95% CI: 0.18, 0.87), respectively] (**Figure 2**). Additional adjustment for self-reported coffee intake did not alter the testosterone findings and only slightly attenuated the results for serum caffeine and paraxanthine and anovulation [aRRs: 0.41 (95% CI: 0.18, 0.98) and 0.42 (95% CI: 0.19, 0.91), respectively]. No significant associations were shown for serum theobromine and anovulation. Similarly, although interactions between race, smoking, and anovulation were nonsignificant, the strongest associations between increased serum caffeine tertile and reduced risk of anovulation was seen in Asians (**Supplemental Figure 2**).

24HDR caffeine and serum caffeine-biomarker agreement

Pearson correlation coefficients between 24HDR caffeine intake (per mg · kg^{-1} · d^{-1}) and serum biomarkers were 0.43 for caffeine, 0.53 for paraxanthine, and 0.13 for theobromine (all significant at $P < 0.001$). Sensitivity analyses by the time of caffeine intake via the 24HDR indicated a gradual reduction in correlations with serum caffeine and paraxanthine (<13 h: r = 0.65 and 0.58, respectively; 13–18 h, r = 0.30 and 0.32, respectively; >18–23 h, r = 0.29 and 0.39, respectively; and >23–28 h, r = 0.20 and 0.27, respectively).

Overall weighted κ agreements between 24HDR caffeine intake and serum biomarker quartiles were 0.44 (95% CI: 0.40, 0.48) for caffeine, 0.44 (95% CI: 0.41, 0.48) for paraxanthine, and 0.25 for theobromine (95% CI: 0.21, 0.30). Overall, sensitivity values (proportions of women with high serum caffeine and paraxanthine concentrations correctly identified by reported intake) were 71.1% (95% CI: 68.2%, 73.9%) and 68.1% (95% CI: 65.1%, 71.1%), respectively; and specificity values (proportions of women with low serum caffeine and paraxanthine correctly identified by reported intake) were 68.5% (95% CI: 65.4%, 71.5%) and 67.2% (95% CI: 64.2%, 70.2%). The relation between reported caffeine intake and serum concentrations of caffeine differed significantly by race (P = 0.01). White women had a higher adjusted κ between 24HDR intake and serum paraxanthine (0.51) than for caffeine (0.45), whereas Asian women had a lower adjusted κ for serum paraxanthine (0.48) than for caffeine (0.52). Pearson correlation coefficients reflected these same differences in caffeine metabolism by race when restricted to <13 h of reported intake (**Supplemental Table 5**). Although sensitivity and specificity were similar for

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018

TABLE 2
Differences in log serum concentrations of reproductive hormones according to serum caffeine, paraxanthine, and theobromine tertiles (n = 259 women)[1]

	Tertile, μmol/L											
	Serum caffeine				Serum paraxanthine				Serum theobromine			
Log hormone	1 (0.09–0.49)	2 (>0.49–2.04)	3 (>2.04–51.26)	*P*-trend	1 (0.00–0.35)	2 (>0.35–1.10)	3 (>1.10–20.50)	*P*-trend	1 (0.00–0.89)	2 (>0.89–3.27)	3 (>3.27–51.14)	*P*-trend
Estradiol, ρg/mL	86.6 (81.2, 92.4)	84.2 (79.3, 89.4)	82.6 (77.5, 88.0)	0.31	85.0 (80.1, 91.2)	85.0 (79.4, 91.0)	83.3 (78.2, 88.8)	0.56	82.9 (77.8, 88.3)	85.7 (80.7, 90.9)	84.7 (79.7, 90.0)	0.77
Free estradiol, ρg/mL	1.37 (1.29, 1.46)	1.31 (1.24, 1.39)	1.29 (1.21, 1.37)	0.18	1.35 (1.28, 1.42)	1.33 (1.24, 1.41)	1.29 (1.22, 1.37)	0.29	1.32 (1.24, 1.40)	1.33 (1.25, 1.40)	1.32 (1.26, 1.41)	0.84
Testosterone, ng/dL	29.1 (27.9, 30.3)[a]	28.2 (27.1, 29.3)[b]	27.9 (26.7, 29.0)[b]	0.02	28.8 (27.7, 30.0)[a]	27.9 (26.8, 29.1)[b]	28.1 (27.0, 29.3)	0.31	28.2 (27.1, 29.4)	28.4 (27.3, 29.6)	28.4 (27.3, 29.6)	0.82
Free testosterone, ng/dL	0.186 (0.179, 0.194)[a]	0.180 (0.174, 0.187)[b]	0.178 (0.171, 0.185)[b]	0.02	0.184 (0.177, 0.191)	0.179 (0.172, 0.186)	0.179 (0.173, 0.187)	0.26	0.182 (0.175, 0.189)	0.181 (0.175, 0.188)	0.181 (0.174, 0.188)	0.71
Luteal progesterone, ng/mL	1.56 (1.39, 1.75)	1.62 (1.46, 1.81)	1.62 (1.45, 1.82)	0.73	1.52 (1.37, 1.68)	1.77 (1.55, 2.01)	1.60 (1.43, 1.79)	0.92	1.53 (1.37, 1.72)	1.63 (1.47, 1.82)	1.63 (1.46, 1.82)	0.52
FSH, mIU/mL	5.52 (5.25, 5.80)	5.31 (5.07, 5.56)	5.24 (4.99, 5.50)	0.18	5.49 (5.24, 5.74)	5.20 (4.94, 5.48)	5.30 (5.05, 5.56)	0.51	5.43 (5.17, 5.69)	5.37 (5.13, 5.62)	5.28 (5.04, 5.53)	0.32
LH, ng/mL	6.89 (6.46, 7.36)[a]	6.13 (5.76, 6.51)[b]	6.09 (5.71, 6.50)[b]	0.04	6.68 (6.30, 7.09)[a]	5.97 (5.56, 6.42)[b]	6.23 (5.85, 6.65)	0.35	6.46 (6.05, 6.89)	6.54 (6.16, 6.95)	6.10 (5.73, 6.48)	0.08
Testosterone:estradiol ratio	0.77 (0.76, 0.79)	0.77 (0.76, 0.78)	0.77 (0.76, 0.79)	0.91	0.78 (0.76, 0.79)	0.77 (0.75, 0.78)	0.77 (0.76, 0.79)	0.74	0.78 (0.76, 0.79)	0.77 (0.76, 0.79)	0.77 (0.76, 0.78)	0.48

[1]All values are mean βs; 95% CIs in parentheses. Analyses were performed with the use of linear mixed-effects models on the log scale of hormones with adjustment for age, race, percentage of body fat, depression, time-varying daily vigorous exercise, perceived stress, Mediterranean diet score, and total energy and alcohol intakes. Mean values that do not share a common superscript letter were significantly different at $P < 0.05$ on the basis of Tukey's test. Overall, there were ≥3650 measurements (90%) for each reproductive hormone collected from ≤16 clinic visits across 2 menstrual cycles (n = 250 women) or ≤8 clinic visits for 1 menstrual cycle (n = 9 women). Note: 1.00 μmol serum caffeine/L = 194.2 ng/mL; 1.00 μmol serum paraxanthine/L = 180.2 ng/mL; and 1.00 μmol serum theobromine/L = 180.2 ng/mL. FSH, follicle-stimulating hormone; LH, luteinizing hormone.

whites and Asians, blacks had lower sensitivity and higher specificity relative to those of other races (**Supplemental Table 6**). A significant interaction was also shown between the 24HDR and serum caffeine by smoking status (P = 0.001). Stratification by smoking status indicated that smokers metabolized caffeine more quickly than did nonsmokers with a relatively higher adjusted κ for serum paraxanthine than for serum caffeine for smokers (0.56 compared with 0.46, respectively) compared with those for nonsmokers (0.50 compared with 0.48, respectively). No significant interactions were shown between caffeine biomarkers and menstrual cycle phase.

DISCUSSION

In answering the question of whether the actual biological dose of caffeine affects reproductive function, we showed that serum concentrations of caffeine and paraxanthine were significantly associated with reduced total testosterone, free testosterone, and LH concentrations and reduced risk of anovulatory episodes irrespective of the caffeinated beverage source, which suggested that moderate caffeine intake may be associated with improved menstrual cycle function. Although we showed no significant effect modification by race, the strongest findings, both in magnitude and significance, for reduced testosterone concentrations and improved ovulatory function with increasing serum caffeine were shown in Asians. In answering the question of whether self-reported caffeine intake reflects the biological dose, we showed near-borderline agreement and correlation for what is considered a useful measure between dietary intake and biomarkers (κ = 0.40, r = 0.50), which suggested limited validity (12, 26). The relatively higher agreement for paraxanthine than for caffeine in whites and smokers supported the more-rapid caffeine metabolism of these groups as previously reported (5, 27, 28). Because of our results, we advise that serum caffeine biomarkers, alone or in combination with reported intake, be used when possible to reduce the misclassification of caffeine exposure when assessing the association between caffeine exposure and women's reproductive function.

Our finding of significantly reduced total testosterone, free testosterone, and LH after adjustment for relevant confounding factors has been supported by previous research. A recent randomized controlled trial in 42 healthy, nonsmoking adults (24 women), who, after a 2-wk abstention from caffeine, were randomly assigned to consume five 6-oz cups of caffeinated coffee, decaffeinated coffee, or water (controls), showed that both caffeinated coffee and decaffeinated coffee were significantly associated with decreased free testosterone relative to that of controls (29). Caffeinated coffee and decaffeinated coffee also showed modest but nonsignificant, associations with reduced free estradiol with no indication of significant differences in the ratio of testosterone to estradiol, which is a potential marker for aromatase activity (29). We previously reported that self-reported 24HDR caffeine intake ≥200 compared with <200 mg/d was significantly associated with decreased total estradiol, free estradiol, and LH in whites and increased total estradiol, free estradiol, and LH in Asians after adjustment for relevant confounders (15). In the current study, serum caffeine was more-significantly associated with decreased total testosterone and free testosterone, notably in Asians. Although previous studies have shown differences by race and genotype in both drug and

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018

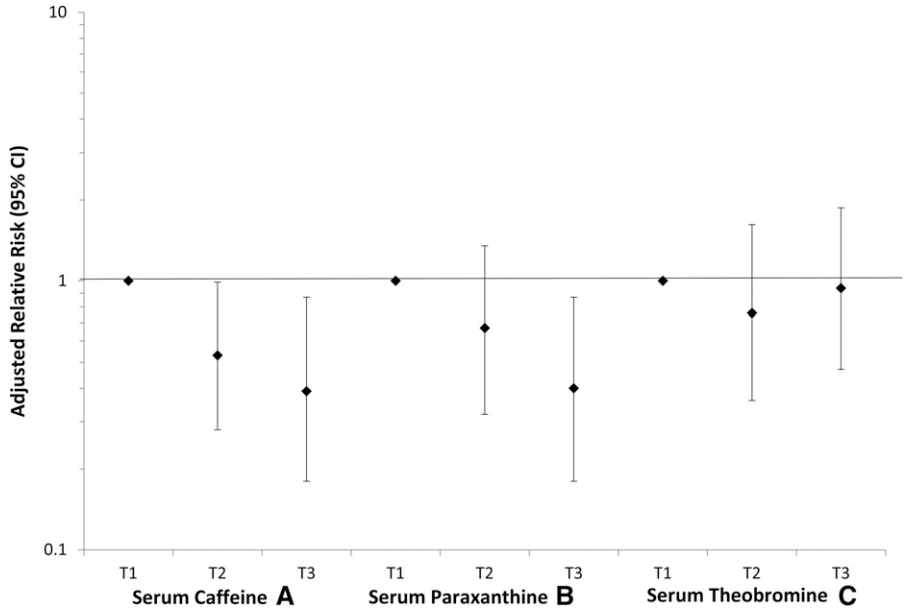

FIGURE 2 Adjusted RRs (95% CIs) of anovulation according to serum caffeine (A), paraxanthine (B), and theobromine (C) tertiles (n = 259 women). Analyses were performed with the use of generalized linear mixed models and adjusted for age, race, percentage of body fat, daily vigorous exercise, perceived stress, depression, Mediterranean diet score, and total energy and alcohol intakes (all continuous except for race, which was categorized as white, black, Asian, or other). A total of 259 women were followed for ≤1 (n = 9) or 2 (n = 250) menstrual cycles. Anovulation was defined as any cycle with a peak progesterone concentration ≤5 ng/mL and no observed serum luteinizing hormone peak at the midluteal or late-luteal phase visits (n = 42 of 509 cycles; 8.3%); 28 women had 1 anovulatory cycle, and 7 women had 2 anovulatory cycles. T, tertile.

steroid metabolism (27–32), we did not have genotype data for this study, and self-reported race represents a complex interplay between many factors beyond genetics including social, lifestyle, and environmental factors. Future studies that measure polymorphisms of the cytochrome P450 enzymes that are responsible for caffeine and steroid synthesis are needed to tease apart the potential interplay between race, caffeine, and reproductive function.

As regards ovulatory function, we showed caffeine and its chief metabolite paraxanthine were associated with higher luteal progesterone and significant reduced risk of sporadic anovulation. These results were not inconsistent with our previous research whereby 24HDR caffeine intake ≥200 compared with <200 mg/d was associated with a reduced (albeit nonsignificant) odds for sporadic anovulation (OR: 0.82; 95% CI: 0.31, 2.21). After we accounted for brand and brewing variations in caffeine doses and differences in caffeine metabolism by assessing actual biological dose via caffeine biomarkers, we may have increased the precision to identify a significant association between caffeine and ovulatory function. Additional adjustment for self-reported coffee intake did not alter the reproductive hormone findings and only slightly attenuated the results for serum caffeine and paraxanthine and anovulation, which indicated that the caffeine, rather than some other component, in coffee was the explanatory factor. Our finding has been supported by a previous study that was conducted to assess the effects of self-reported caffeine and coffee intakes on luteal progesterone in a cohort of, on average, higher caffeine consumers, which showed significantly increased geometric mean concentrations of luteal progesterone (ng/mL) with increasing caffeine tertiles (≤70, >70–190, >190–371, and >371 mg/d; P-trend = 0.03) but not with coffee intake (30). Although both this previous finding and our results support the hypothesis that increased caffeine intake leads to decreased testosterone with a consequent decreased risk of sporadic anovulation (33), whether there is some other component in coffee, such as polyphenols, that is associated with increased progesterone (34) or some such combination of components warrants future investigation.

We previously reported that 24HDR self-reported caffeine and caffeinated beverage intakes (coffee, tea, and soda) do not vary over the cycle (35). In this study, we report no significant variations in serum caffeine and caffeine metabolites across the menstrual cycle and no significant interaction between 24HDR and serum caffeine by cycle phase. Taken together, we conclude that caffeine metabolism does not appear to vary across the menstrual cycle. Although, to our knowledge, our study is the first observational trial to assess caffeine metabolism across the menstrual cycle, our findings are in agreement with the majority of experimental studies (11, 14).

The BioCycle Study has several strengths including multiple measures of hormone and serum caffeine and caffeine metabolite assessment over 2 menstrual cycles and multiple measures of not only caffeine and caffeinated beverage intake but also of important dietary and lifestyle factors. Our relatively diverse study population (~60% white, 20% black, 14% Asian, and 7% other race) allowed us to be able to detect interracial differences in either the reporting of caffeine-containing foods and beverages or caffeine metabolism, which shed light on our previous findings of differences in the effects of reported caffeine intake and reproductive hormones (15). In addition, we were able to assess ovulatory function with the use of repeated measures of luteal progesterone after making sure that our assessments were appropriately timed and included adequate serum draws through the luteal phase via the identification of a midcycle LH serum surge (22).

Nevertheless, our study has several limitations including the relatively limited caffeine intake by our study population (US

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018

premenopausal women consume an average of 166 mg total caffeine/d (1) compared with that by our population who reported consuming an average of 91 mg/d). In addition, although our ability to assess differences between the time of caffeine intake and serum caffeine and metabolite concentrations was an improvement over previous observational studies in premenopausal women (36), the relatively rapid metabolism of caffeine with an average half-life of 5–6 h resulted in relatively low serum concentrations than in studies that have assessed nonfasting serum biomarkers (36). Note that our study relied on observational data in which a multitude of analyses have been made, 5% of which could have been significant purely by chance. However, we formally adjusted for multiple comparisons to reduce the type I error rate via Tukey's test and further report 95% CIs for all of our estimates, and thus, the reader can evaluate the strength and precision of our significant findings. Finally, we purposely limited our study population to healthy female volunteers with no previous history of reproductive disorders or chronic disease so as to better understand the subtle associations between diet and oxidative stress and reproductive hormones. Consequently, the generalizability of our findings to women with evident reproductive dysfunction may be limited.

In conclusion, we showed little variation of serum caffeine and metabolites across the menstrual cycle and no evidence for differences in caffeine clearance between phases in line with previous research (11, 14). These results suggest that the normal hormonal fluctuations observed in eumenorrheic healthy women do not appear to alter caffeine pharmacokinetics. Although κ estimates for 24HDR and serum caffeine and paraxanthine quartiles were within the range of agreement for most duplicate measures of nutritional status, our significant findings for lower risks of sporadic anovulation in the highest tertiles of serum caffeine and paraxanthine were not shown in our previously published results that looked at 24HDR caffeine ($\geq$100, $\geq$200, $\geq$300, and $\geq$400 mg/d) or coffee (1-cup increments) intake via various cutoffs (15), thereby suggesting an increased precision with the serum caffeine biomarkers. Our findings in conjunction with those of other authors (29) fail to support the mechanistic hypothesis that caffeine is an aromatase inhibitor (30). Rather, caffeine may help to keep testosterone concentrations in check, thereby lowering the risk of anovulatory episodes (33). Although more good-quality, adequately powered observational studies that assess the relation between caffeine biomarkers, alone or in combination with self-reported intake, and menstrual cycle are needed before definitive conclusions can be made, results from the current study suggest a possible protective effect of moderate caffeine intake on ovulatory function in healthy, premenopausal women.

The authors' responsibilities were as follows—EFS and JW-W: designed and conducted the research; KCS, NJP, and SLM: performed the statistical analysis; KCS and SLM: wrote the manuscript; KCS: had primary responsibility for the final content of the manuscript; and all authors: interpreted the data, assisted with manuscript revisions, and read and approved the final manuscript. None of the authors reported a conflict of interest related to the study.

REFERENCES

1. Frary CD, Johnson RK, Wang MQ. Food sources and intakes of caffeine in the diets of persons in the United States. J Am Diet Assoc 2005;105:110–3.
2. McCusker RR, Goldberger BA, Cone EJ. Caffeine content of specialty coffees. J Anal Toxicol 2003;27:520–2.
3. Carl J, Hill DA. Preconception counseling: make it part of the annual exam. J Fam Pract 2009;58:307–14.
4. Peck JD, Leviton A, Cowan LD. A review of the epidemiologic evidence concerning the reproductive health effects of caffeine consumption: a 2000–2009 update. Food Chem Toxicol 2010;48:2549–76.
5. Klebanoff MA, Levine RJ, Dersimonian R, Clemens JD, Wilkins DG. Serum caffeine and paraxanthine as markers for reported caffeine intake in pregnancy. Ann Epidemiol 1998;8:107–11.
6. Kotsopoulos J, Eliassen AH, Missmer SA, Hankinson SE, Tworoger SS. Relationship between caffeine intake and plasma sex hormone concentrations in premenopausal and postmenopausal women. Cancer 2009;115:2765–74.
7. Nagata C, Kabuto M, Shimizu H. Association of coffee, green tea, and caffeine intakes with serum concentrations of estradiol and sex hormone-binding globulin in premenopausal Japanese women. Nutr Cancer 1998; 30:21–4.
8. Lucero J, Harlow BL, Barbieri RL, Sluss P, Cramer DW. Early follicular phase hormone levels in relation to patterns of alcohol, tobacco, and coffee use. Fertil Steril 2001;76:723–9.
9. Chavarro JE, Rich-Edwards JW, Rosner BA, Willett WC. Caffeinated and alcoholic beverage intake in relation to ovulatory disorder infertility. Epidemiology 2009;20:374–81.
10. Fenster L, Quale C, Waller K, Windham GC, Elkin EP, Benowitz N, Swan SH. Caffeine consumption and menstrual function. Am J Epidemiol 1999;149:550–7.
11. Zaigler M, Rietbrock S, Szymanski J, Dericks-Tan JS, Staib AH, Fuhr U. Variation of CYP1A2-dependent caffeine metabolism during menstrual cycle in healthy women. Int J Clin Pharmacol Ther 2000;38:235–44.
12. Willett W. Nutritional epidemiology. 3rd ed. New York: Oxford University Press; 2013.
13. Lane JD, Steege JF, Rupp SL, Kuhn CM. Menstrual cycle effects on caffeine elimination in the human female. Eur J Clin Pharmacol 1992; 43:543–6.
14. Kamimori GH, Joubert A, Otterstetter R, Santaromana M, Eddington ND. The effect of the menstrual cycle on the pharmacokinetics of caffeine in normal, healthy eumenorrheic females. Eur J Clin Pharmacol 1999;55:445–9.
15. Schliep KC, Schisterman EF, Mumford SL, Pollack AZ, Zhang C, Ye A, Stanford JB, Hammoud AO, Porucznik CA, Wactawski-Wende J. Caffeinated beverage intake and reproductive hormones among premenopausal women in the BioCycle Study. Am J Clin Nutr 2012;95: 488–97.
16. Wactawski-Wende J, Schisterman EF, Hovey KM, Howards PP, Browne RW, Hediger M, Liu A, Trevisan M. BioCycle Study: design of the longitudinal study of the oxidative stress and hormone variation during the menstrual cycle. Paediatr Perinat Epidemiol 2009;23:171–84.
17. Kalow W, Tang BK. Use of caffeine metabolite ratios to explore CYP1A2 and xanthine oxidase activities. Clin Pharmacol Ther 1991;50:508–19.
18. Whitcomb BW, Perkins NJ, Albert PS, Schisterman EF. Treatment of batch in the detection, calibration, and quantification of immunoassays in large-scale epidemiologic studies. Epidemiology 2010;21:S44–50.
19. Howards PP, Schisterman EF, Wactawski-Wende J, Reschke JE, Frazer AA, Hovey KM. Timing clinic visits to phases of the menstrual cycle by using a fertility monitor: the BioCycle Study. Am J Epidemiol 2009; 169:105–12.
20. Sartorius G, Ly LP, Sikaris K, McLachlan R, Handelsman DJ. Predictive accuracy and sources of variability in calculated free testosterone estimates. Ann Clin Biochem 2009;46:137–43.
21. Södergård R, Backstrom T, Shanbhag V, Carstensen H. Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. J Steroid Biochem 1982;16:801–10.
22. Lynch KE, Mumford SL, Schliep KC, Whitcomb BW, Zarek SM, Pollack AZ, Bertone-Johnson ER, Danaher M, Wactawski-Wende J, Gaskins A, et al. Assessment of anovulation in eumenorrheic women: comparison of ovulation detection algorithms. Fertil Steril 2014;102: 511–8.e2.
23. Prasad A, Schisterman EF, Schliep KC, Ahrens KA, Sjaarda LA, Perkins NJ, Matyas R, Wactawski-Wende J, Mumford SL. Depressive symptoms and their relationship with endogenous reproductive hormones and sporadic anovulation in premenopausal women. Ann Epidemiol 2014;24:920–4.

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018

24. Gaskins AJ, Rovner AJ, Mumford SL, Yeung E, Browne RW, Trevisan M, Perkins NJ, Wactawski-Wende J, Schisterman EF, BioCycle Study Group. Adherence to a Mediterranean diet and plasma concentrations of lipid peroxidation in premenopausal women. Am J Clin Nutr 2010; 92:1461–7.
25. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther 2005;85:257–68.
26. Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. Am J Clin Nutr 1997;65:1130S–7S.
27. Gunes A, Dahl ML. Variation in CYP1A2 activity and its clinical implications: influence of environmental factors and genetic polymorphisms. Pharmacogenomics 2008;9:625–37.
28. Mizutani T. PM frequencies of major CYPs in Asians and Caucasians. Drug Metab Rev 2003;35:99–106.
29. Wedick NM, Brennan AM, Sun Q, Hu FB, Mantzoros CS, van Dam RM. Effects of caffeinated and decaffeinated coffee on biological risk factors for type 2 diabetes: a randomized controlled trial. Nutr J 2011;10:93.
30. Kotsopoulos J, Vitonis AF, Terry KL, De Vivo I, Cramer DW, Hankinson SE, Tworoger SS. Coffee intake, variants in genes involved in caffeine metabolism, and the risk of epithelial ovarian cancer. Cancer Causes Control 2009;20:335–44.
31. Adlercreutz H, Gorbach SL, Goldin BR, Woods MN, Dwyer JT, Hamalainen E. Estrogen metabolism and excretion in Oriental and Caucasian women. J Natl Cancer Inst 1994;86:1076–82.
32. MacMahon B, Cole P, Brown JB, Aoki K, Lin TM, Morgan RW, Woo NC. Urine oestrogen profiles of Asian and North American women. Int J Cancer 1974;14:161–7.
33. Sjaarda LA, Mumford SL, Kissell K, Schliep KC, Hammoud AO, Perkins NJ, Weck J, Wactawski-Wende J, Schisterman EF. Increased androgen, anti-Mullerian hormone, and sporadic anovulation in healthy, eumenorrheic women: a mild PCOS-like phenotype? J Clin Endocrinol Metab 2014;99:2208–16.
34. Lin CJ, Cheng LC, Lin TC, Wang CJ, Li LA. Assessment of the potential of polyphenols as a CYP17 inhibitor free of adverse corticosteroid elevation. Biochem Pharmacol 2014;90:288–96.
35. Schliep KC, Schisterman EF, Mumford SL, Perkins NJ, Ye A, Pollack AZ, Zhang C, Porucznik CA, VanDerslice JA, Stanford JB, et al. Validation of different instruments for caffeine measurement among premenopausal women in the BioCycle study. Am J Epidemiol 2013; 177:690–9.
36. Klebanoff MA, Levine RJ, DerSimonian R, Clemens JD, Wilkins DG. Maternal serum paraxanthine, a caffeine metabolite, and the risk of spontaneous abortion. N Engl J Med 1999;341:1639–44.

Downloaded from https://academic.oup.com/ajcn/article-abstract/104/1/155/4569660 by guest on 03 December 2018