Practice of Epidemiology

Validation of Different Instruments for Caffeine Measurement Among Premenopausal Women in the BioCycle Study

Karen C. Schliep, Enrique F. Schisterman*, Sunni L. Mumford, Neil J. Perkins, Aijun Ye, Anna Z. Pollack, Cuilin Zhang, Christina A. Porucznik, James A. VanDerslice, Joseph B. Stanford, and Jean Wactawski-Wende

* Correspondence to Dr. Enrique F. Schisterman, Epidemiology Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, 6100 Executive Boulevard, 7B03M, Rockville, MD 20852 (e-mail: schistee@mail.nih.gov).

Effects of caffeine on women's health are inconclusive, in part because of inadequate exposure assessment. In this study we determined 1) validity of a food frequency questionnaire compared with multiple 24-hour dietary recalls (24HDRs) for measuring monthly caffeine and caffeinated beverage intakes; and 2) validity of the 24HDR compared with the prior day's diary record for measuring daily caffeinated coffee intake. BioCycle Study (2005–2007) participants, women ($n = 259$) aged 18–44 years from western New York State, were followed for 2 menstrual cycles. Participants completed a food frequency questionnaire at the end of each cycle, four 24HDRs per cycle, and daily diaries. Caffeine intakes reported for the food frequency questionnaires were greater than those reported for the 24HDRs (mean = 114.1 vs. 92.6mg/day, $P = 0.01$) but showed high correlation ($r = 0.73$, $P < 0.001$) and moderate agreement (κ = 0.51, 95% confidence interval: 0.43, 0.57). Women reported less caffeinated coffee intake in their 24HDRs compared with their corresponding diary days (mean = 0.51 vs. 0.80 cups/day, $P < 0.001$) (1 cup = 237 mL). Although caffeine and coffee exposures were highly correlated, absolute intakes differed significantly between measurement tools. These results highlight the importance of considering potential misclassification of caffeine exposure.

beverages; caffeine; diet; mental recall; nutrition assessment; questionnaires; validation studies; women

Abbreviations: FFQ, food frequency questionnaire; 24HDR, 24-hour dietary recall.

Caffeine has received a great deal of attention regarding its health effects on premenopausal women (1–3). Coffee, tea, and soda contain components in addition to caffeine that may affect health, highlighting the importance of beverage type (4, 5). Studies of the health effects of caffeine and caffeinated beverages have been inconclusive, partly because of inadequate exposure assessment (2, 3).

Measuring caffeine intake is difficult because caffeine is present in a variety of sources (6). Additionally, caffeine exposure can vary depending on the brand, serving size, and method of preparation of the food or beverage (7). Retrospective assessment of caffeine intake that is recalled at a single time point may be prone to error if it fails to account for exposure fluctuations (2), and prospective assessment may lack precision if it fails to capture caffeine source and serving size (8). Such exposure misclassification may bias effect estimates toward or away from the null depending on the magnitude and direction of the exposure errors (2), highlighting the importance of assessing the validity of common methods for measuring caffeine consumption.

Caffeine intake tends to increase with age (9, 10), and both the metabolism of caffeine and, potentially, caffeine intake behavior are affected by reproductive hormones during the menstrual cycle (11). Therefore, arriving at a valid method for assessing caffeine exposure in this population is essential for a clear understanding of the effect of

caffeine on women's reproductive health. Most studies assessing caffeine among nonpregnant, premenopausal women use self-administered, semiquantitative food frequency questionnaires (FFQs) (5, 10, 12–14). Diet records and recalls are generally considered to be the "gold standard" for dietary assessment and thus are often used as the reference when assessing the relationship between reported intakes from the FFQ and true usual intakes (15). Various versions of the FFQ have been validated for caffeine intake among non-American, older women in 2 previous studies (mean ages of subjects: 54 and 58 years) (15, 16) and for caffeinated beverage intake (i.e., of coffee, tea, and soda) among pre- to perimenopausal American women (aged 34–59 years, uniformly distributed) (17). No study to date, however, has assessed the validity of the FFQ for both caffeine and caffeinated beverage intakes for a younger cohort of American women of reproductive age by using appropriate statistical methods. Validation studies depending on correlation analyses alone are inadequate because correlation measures the strength of the linear relationship, not the agreement, and correlations depend on the range of the true quantity within the sample (18).

Our primary objective was to assess the validity of 1) the FFQ compared with multiple 24-hour dietary recalls (24HDRs) for measuring monthly caffeine and caffeinated beverage intakes; and 2) the 24HDR compared with the corresponding day's daily diary record for measuring daily caffeinated coffee intake. Our secondary objective was to assess the variability of caffeine consumption patterns by comparing 1) caffeine and caffeinated beverage intakes for eight 24HDRs and (for caffeinated coffee) daily diaries captured over 2 menstrual cycles; and 2) caffeine and caffeinated beverage intakes as reported on the FFQ completed at baseline (capturing intakes during the previous 6 months) with the FFQ completed at the end of each menstrual cycle (capturing intakes during the previous menstrual cycle).

MATERIALS AND METHODS

Study population

The BioCycle Study (2005–2007) included women aged 18–44 years from western New York State who were enrolled for 1 ($n = 9$) or 2 ($n = 250$) menstrual cycles. The study population, materials, and methods have been previously described in detail (19). In summary, eligible women had a body mass index (measured as weight (kg)/height (m)^2) of between 18 and 35 and no history of chronic disease. Women who were currently pregnant, had been pregnant in the last 6 months, or were planning to attempt to conceive in the next 3 months were ineligible. Physical measures were obtained in the clinic by using standardized protocols, and sociodemographic and lifestyle information was collected by using validated questionnaires (19). The Health Sciences Institutional Review Board at the University at Buffalo approved the study and served as the institutional review board designated by the National Institutes of Health under a reliance agreement.

Dietary assessment

Twenty-four hour dietary recall. Participants completed a 24HDR at the clinic after the collection of a fasting blood specimen during the visits corresponding to menstruation, the midfollicular phase, ovulation, and the midluteal phase. Study visits were scheduled to occur during these key phases of the menstrual cycle by using an algorithm accounting for each woman's self-reported cycle length. Fertility monitors (Clearblue Easy Fertility Monitor; Inverness Medical Innovations, Inc., Waltham, Massachusetts) measured estrone-3-glucuronide and leuteinizing hormone in urine starting on calendar day 6 after menses and continuing for 10–20 days. The monitor algorithm determines "peak fertility" on the basis of predetermined cutpoints during the first cycle, but in subsequent cycles it adjusts the cutpoint criteria according to the woman's specific hormone levels. Monitor indications of low, high, and peak fertility were used to time midcycle visits.

Information regarding food and beverage intakes was collected by using a standardized, multiple-pass approach of interview methodology. Nutrient intakes were calculated by using the Nutrition Data System for Research, 2005 version (Nutrition Coordinating Center, Minneapolis, Minnesota). This program calculates caffeine intake on the basis of consumption-weighted averages of US Department of Agriculture values, computing the nutrients (e.g., caffeine in mg/day) and the food and beverage components (e.g., unsweetened coffee in cups/day) (1 cup = 237 mL) from the 24HDR assessments. The Nutrition Data System for Research assumes the following caffeine contents: 94.7 mg per cup of caffeinated coffee; 48.8 mg per cup of reduced caffeine coffee; 62.8 mg per fluid ounce of espresso (1 fluid ounce = 30 mL); 47.4 mg per cup of brewed black or green tea; 2.4 mg per cup of decaffeinated tea (black or green) or decaffeinated coffee; 24.8 mg per cup of caffeinated cola soda; 36.8 mg per cup of highly caffeinated soda (e.g., Mountain Dew) (PepsiCo, Inc., Purchase, New York); 17 mg per ounce of dark chocolate (1 ounce = 28 g); and 6 mg per ounce of milk chocolate. Ninety-six percent of participants completed at least three 24HDRs in both of their cycles and 73% completed all eight 24HRDs.

Food frequency questionnaire. Nutrient data were collected by using the Nutrition Assessment Shared Resource FFQ-GSEL (Fred Hutchinson Cancer Research Center, Seattle, Washington) (20). Participants completed the FFQ up to 3 times at the clinic: once at baseline (FFQ-B) to capture the usual intake for the previous 6 months and once during the late luteal phase of each cycle (FFQ-1 and FFQ-2) to determine the usual intake in the month of each cycle. Participants reported on the frequency of consumption (e.g., ranging from "never or less than once per month" to "6+ per day") and portion size (e.g., small, medium, or large, with the medium serving size of coffee described as 1 cup) of approximately 120 items, including 5 caffeinated beverages and chocolate. These items were grouped as follows: "latte, cappuccino, mocha, or hot chocolate" (hereafter referred to as "coffee drinks/cocoa"); "coffee (not lattes or mochas)"; "tea (all types)"; "diet soft drinks"; "regular soft drinks"; and "chocolate candy bars and toffee." The Nutrition Assessment

Shared Resource provided estimates of daily intakes of nutrients including caffeine by using the same Nutrition Data System for Research software version as we used for the 24HDR. Ninety-nine percent of the participants completed at least 1 FFQ and 86% completed all 3 FFQs (FFQ-B, FFQ-1, and FFQ-2).

Daily diary. Participants used daily diary forms to record their daily caffeinated coffee intake and other lifestyle and health items, including the number of cigarettes smoked. No other caffeinated foods or beverages were recorded on the daily diary. At baseline, study staff instructed each participant to begin completing her daily diaries on the first day of her next menstrual period and to continue through the next 2 menstrual cycles. Participants recorded the number of cups of caffeinated coffee (hot or iced, instant or brewed) they consumed. Ninety-seven percent of participants completed at least 75% of the daily diaries in at least 1 cycle; 71% of participants completed 100% of the daily diaries in at least 1 cycle.

Statistical analysis

Validity of caffeine and caffeinated beverage intakes. Although the 24HDR has been shown to outperform the FFQ, daily diaries are thought to be the most accurate self-report method given their prospective versus retrospective nature (21). The BioCycle Study allowed us to compare total caffeine and caffeinated beverage intakes as reported in the FFQ (test method) with those reported in up to eight 24HDRs (reference method). We were also able to compare caffeinated coffee intake as reported in the 24HDR (test method) with that reported in the daily diary (reference method). Because the FFQ-GSEL does not distinguish between caffeinated and decaffeinated coffee, we could not directly compare caffeinated coffee intake reported in the FFQ with that reported in the daily diary.

Descriptive statistics were calculated by including sociodemographic characteristics of the participants. To evaluate the validity of the FFQ for assessing monthly caffeine and caffeinated beverage intakes, we summed and averaged data from the FFQ-1 and the FFQ-2 and compared that mean value with the mean value of the eight 24HDRs; we additionally compared the four 24HDRs per cycle with their corresponding FFQs. Caffeine and caffeinated beverage intakes from the 24HDRs and the FFQs were not normally distributed; therefore, we used nonparametric analysis techniques. We report arithmetic means (standard deviations) for comparison with other studies, in addition to geometric means (standard deviations) and medians (interquartile ranges) to reflect our nonnormally distributed data. To determine the validity of the FFQ compared with the 24HDR, we included in the analyses women who completed either the FFQ-1 or FFQ-2 and at least 75% of their 24HDRs for the corresponding cycle (*n* = 249). To validate the 24HDR compared with the daily diary, we included women who completed at least 75% of their daily diaries and 24HDRs in at least 1 of their cycles (*n* = 251).

Wilcoxon's matched-pairs signed-ranks tests were used to determine differences between the mean ranks. Pearson's product-moment correlation coefficients on log-transformed values described the associations between the FFQs and the 24HDRs. We also calculated deattenuated Pearson's product-moment correlation coefficients in which the within-woman variations were divided by the between-woman variations to quantify the variance ratios of the 24HDRs (22).

To visualize agreement between the FFQ and the 24HDR for caffeine and caffeinated beverage intakes, we constructed Bland-Altman plots by using the mean values of FFQ-1 and FFQ-2 and the eight 24HDRs. We present the plots on the original scale with back-transformed limits of agreement (23). To evaluate the FFQ's ability to assign women to the same categories of intake as the 24HDR, we classified women into 4 categories: nonconsumers of caffeine and tertiles of caffeine and caffeinated beverage intakes among caffeine consumers based on the distribution of data from both the FFQ and the 24HDR (24, 25). We performed cross-classification analyses and compared percent agreement and weighted κ coefficients calculated with a linear set of weights by using Landis and Koch's guidelines for interpreting κ coefficients (26). The recommended level of daily caffeine intake (≤200 mg/day) for women planning to conceive (27) was used as the threshold value to estimate specificity, sensitivity, and positive and negative predictive values of the FFQ, whereby intakes within the recommended range were defined as positive. Sensitivity analysis was performed by using a cutpoint of 300 mg/day of caffeine intake.

To evaluate the validity of the 24HDR for assessing daily caffeinated coffee intake, we used the above analyses to compare the mean caffeine intake as reported on the 24HDR with the previous day's daily diary and classified women into 4 categories: nonconsumers of caffeine and tertiles based on the distribution of data from both the 24HDRs and the daily diaries (24, 25) to evaluate the 24HDR's ability to assign women to the same categories of intake as the daily diary. We chose a relevant cutpoint of 1 cup/day (14) to estimate specificity, sensitivity, and positive and negative predictive values of the 24HDR, whereby intakes of <1 cup/day were defined as positive.

Variability of intakes of caffeine and caffeinated beverages. To determine the variability of caffeine and caffeinated beverage intakes over the study period as reported in the FFQ, we repeated the above analyses to assess the agreement between the FFQ-1 and the FFQ-2 (with the exception of deattenuating the correlation coefficients, which are deemed unnecessary for reproducibility studies) (24) by restricting analyses to women who completed both the FFQ-1 and the FFQ-2 (*n* = 224). We compared the FFQ-B with the mean of the FFQ-1 and the FFQ-2 to account for changes in consumption while under observation, and we restricted analyses to women who completed all 3 FFQs (*n* = 222).

To determine the variability of caffeine and caffeinated beverage intakes as reported in the 24HDRs and the daily diaries, we used repeated measures analyses with random intercepts and restricted analyses to women who completed at least 75% of their 24HDRs (*n* = 258) and daily diaries (*n* = 251) for at least 1 of their cycles. *P* values are 2-tailed with significance set at *P* < 0.05. Analyses were performed in SAS, version 9.2, software (SAS Institute, Inc., Cary, North Carolina).

RESULTS

Population characteristics

Participants included in the validity study (*n* = 249) were relatively young, with a mean age of 27.5 (standard deviation, 8.3) years; of normal weight, with a mean body mass index of 24.1 (standard deviation, 3.8); predominately white (59.2%); currently nonsmokers (defined as no current cigarette use as recorded in their daily diaries) (95.6%); and nulligravidas (69.1%). Demographics of the women included in the variability analyses as assessed by the FFQ (*n* = 224) were similar to those of subjects in the validity study.

Validity of caffeine and caffeinated beverage intakes

According to the FFQ-B self-reports, 58% of subjects consumed coffee, 72% consumed tea, 64% consumed coffee drinks/cocoa, and 78% consumed soda. Similar patterns were seen for the FFQ-1 and FFQ-2, with 60% of subjects consuming coffee, 72% consuming tea, 63% consuming coffee drinks/cocoa, and 81% consuming soda. Average beverage consumption reported on the 24HDRs was less than that reported on the FFQs, with 49% of subjects consuming coffee, 64% consuming tea, 21% consuming coffee drinks/cocoa, and 71% consuming soda.

Compared with the 24HDR, the FFQ overestimated usual daily intakes of caffeine (mean = 114.1 mg/day vs. 92.6 mg/day; geometric mean = 48.9 vs. 41.4; *P* = 0.01), coffee (mean = 0.76 cups/day vs. 0.51 cups/day; geometric mean = 0.11 vs. 0.08; *P* < 0.001), and coffee drinks/cocoa (mean = 0.18 cups/day vs. 0.09 cups/day; geometric mean = 0.05 vs. 0.02; *P* < 0.001) and underestimated usual daily soda intake (mean = 0.41 cups/day vs. 0.57 cups/day; geometric mean = 0.12 vs. 0.16; *P* < 0.001), although the log-transformed caffeine and caffeinated beverage intakes

were all significantly correlated (*P* < 0.001) (Table 1). Despite divergence, the Bland-Altman plots showed acceptable relative limits of agreement (Figure 1). The intra-woman limits of agreement were ±1.14 for caffeine, ±0.94 for coffee, ±1.34 for coffee drinks/cocoa, ±1.45 for tea, and ±1.24 for soda. Differences for intakes of all beverages except coffee followed a normal distribution. Results were similar when we compared the average intakes per cycle as reported for the 24HDRs with the corresponding FFQs (data not shown).

The majority of women (50%–71%) were assigned to the same categories of consumption by both methods except for the consumption of coffee drinks/cocoa, in which 41% were assigned to the same category, 21% to the adjacent category, 24% to the second adjacent category, and 14% to the extreme category; and the consumption of tea, in which 49% were assigned to the same category, 30% to the adjacent category, 12% to the second adjacent category, and 3% to the extreme category (Table 2). Weighted κ values showed substantial agreement for coffee (weighted κ = 0.74); moderate agreement for caffeine, tea, and soda (weighted κ = 0.51, 0.43, and 0.53, respectively); and slight agreement for coffee drinks/cocoa (weighted κ = 0.17). By using recommended daily amounts for caffeine as the threshold value (<200 mg/day), we found that sensitivity of the FFQ was 0.90 and specificity was 0.74. The positive and negative predictive values were 0.96 and 0.45, respectively. Results were similar when we used a 300-mg/day cutpoint, with values of 0.95 for sensitivity, 0.64 for specificity, 0.98 for positive predictive value, and 0.45 for negative predictive value.

High correlation (ρ = 0.77, *P* < 0.001) was found between caffeinated coffee intake as reported on the 24HDR and the corresponding day's daily diary; however, mean intakes differed significantly (mean = 0.51 cups/day vs. 0.80 cups/day; geometric mean = 0.05 vs. 0.08; *P* < 0.001) (Figure 2). Mean differences between the 24HDR and the daily diary were

Table 1. Usual Daily Intakes of Caffeine and Caffeine-Related Beverages Calculated From the FFQ and the 24HDR (*n* = 249) in the BioCycle Study, 2005–2007

	FFQ[a]		24HDR[b]		P Value[c]	Correlation	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		*R*[d]	*R*[e]
Caffeine, mg/day	114.1 (146.1)	68.1 (19.5–147.5)	92.6 (95.1)	59.8 (19.4–140.8)	0.006	0.68	0.73
Coffee, cups/day[f,g]	0.76 (1.34)	0.09 (0.00–1.00)	0.51 (0.75)	0.00 (0.00–0.94)	<0.001	0.91	0.99
Coffee drinks/cocoa, cups/day[h]	0.18 (0.50)	0.05 (0.00–0.15)	0.09 (0.33)	0.00 (0.00–0.00)	<0.001	0.39	0.40
Tea, cups/day	0.38 (0.75)	0.09 (0.00–0.39)	0.36 (0.49)	0.17 (0.00–0.50)	0.38	0.57	0.59
Soda, cups/day	0.41 (0.68)	0.12 (0.03–0.42)	0.57 (0.71)	0.31 (0.00–0.80)	<0.001	0.68	0.71

Abbreviations: FFQ, food frequency questionnaire; 24HDR, 24-hour dietary recall; IQR, interquartile range; SD, standard deviation.

[a] Average of FFQ1 and FFQ2 over 2 cycles.
[b] Average of eight 24HDRs over 2 cycles.
[c] Wilcoxon's matched-pairs signed-ranks test.
[d] Pearson's product-moment correlation coefficient on log-transformed data. All correlations were considered significant at *P* < 0.001.
[e] Pearson's deattenuated product-moment correlation coefficient on log-transformed data.
[f] Coffee is all types but "not lattes or mochas."
[g] 1 cup = 237 mL.
[h] Coffee drinks/cocoa includes "latte, cappuccino, mocha, or hot chocolate."
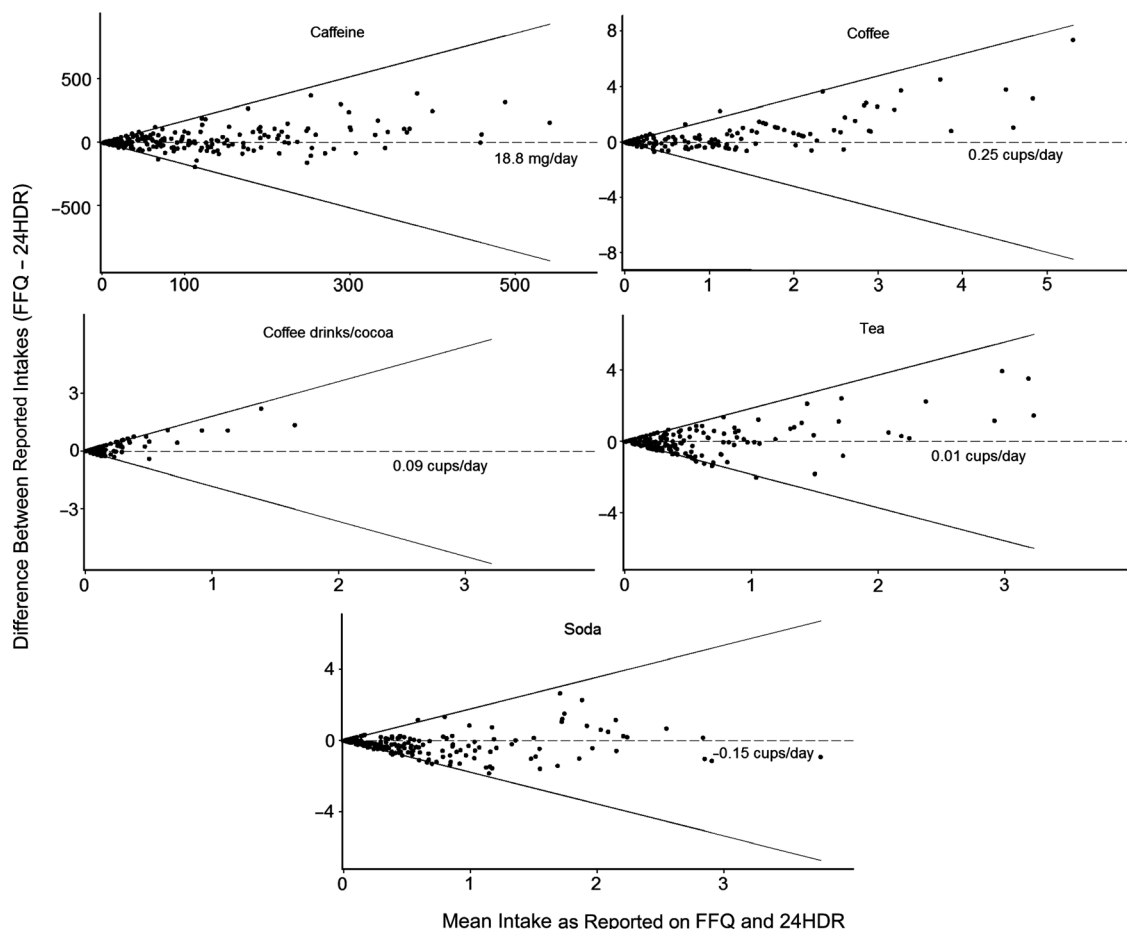
Figure 1.   Bland-Altman plots of difference in caffeine (mg/day) and beverage (cups/day) intakes between the food frequency questionnaire (FFQ) and the 24-hour dietary recall (24HDR) in the BioCycle Study, 2005–2007. The dotted line in each graph represents the mean difference between the FFQ and the 24HDR on the original scale; solid lines represent relative limits of agreement on the logarithmic scale. One cup = 237 mL. Coffee includes all types but "not lattes or mochas." Coffee drinks/cocoa includes "latte, cappuccino, mocha, or hot chocolate."

Table 2.   Cross-classification and κ Coefficient of the FFQ and the 24HDR Categories[a] of Daily Caffeine, Coffee, Coffee Drinks/Cocoa, Tea, and Soda Intakes (*n* = 249) in the BioCycle Study, 2005–2007

	Same Category		Adjacent Category		Extreme Category		Weighted κ	95% CI
	No.	%	No.	%	No.	%		
Caffeine, mg/day	124	50	99	40	5	2	0.51	0.43, 0.58
Coffee, cups/day[b,c]	177	71	63	25	1	0.004	0.74	0.68, 0.80
Coffee drinks/cocoa, cups/day[d]	103	41	53	21	34	14	0.17	0.10, 0.24
Tea, cups/day	123	49	75	30	7	3	0.43	0.35, 0.52
Soda, cups/day	132	53	87	35	1	0.004	0.53	0.46, 0.61

Abbreviations: CI, confidence interval; FFQ, food frequency questionnaire; 24HDR, 24-hour dietary recall.

[a] Caffeine intake is divided into 4 categories on the basis of quartiles; intake of caffeinated beverages is divided into 4 categories: nonconsumers and tertiles.

[b] Coffee is all types but "not lattes or mochas."

[c] 1 cup = 237 mL.

[d] Coffee drinks/cocoa includes "latte, cappuccino, mocha, or hot chocolate."
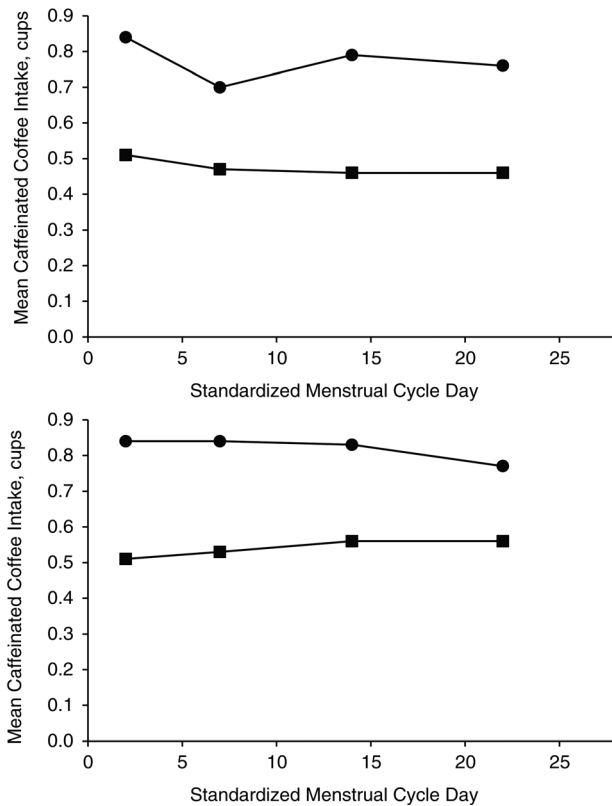
Figure 2. Caffeinated coffee intake reported in the 24-hour dietary recall (24HDR) (squares) at the clinic visit compared with that reported in the daily diary (circles) on the day preceding the clinic visit over 2 menstrual cycles ($n = 258$) (cycle 1, top; cycle 2, bottom) in the BioCycle Study, 2005–2007. The 24HDRs were conducted 4 times per cycle with visits timed via fertility monitors to relevant menstrual cycle phases: day 2 of menstruation, midfollicular phase, day of ovulation, and midluteal phase (approximately corresponding to days 2, 7, 14, and 22 of a 28-day cycle). One cup = 237 mL. Coffee includes all types but "not lattes or mochas." Coffee drinks/cocoa includes "latte, cappuccino, mocha, or hot chocolate."

similar for both cycles. The majority (76%) of women were assigned to the same categories in both methods and 12% were assigned to the adjacent category, 8% to the second adjacent category, and 4% to the extreme category. By using recommended daily amounts for caffeinated coffee as the threshold value (<1 cup/day), we found the sensitivity of the 24HDR to be 0.97 and the specificity to be 0.68. The positive and negative predictive values were 0.81 and 0.94.

Variability of caffeine and caffeinated beverage intakes

Mean daily intakes of caffeine and caffeinated beverages as reported on the FFQ-1 and the FFQ-2 were highly correlated ($R = 0.72–0.94$), as were intakes for 1) FFQ-B and 2) the mean of FFQ-1 and FFQ-2 ($R = 0.76–0.94$) (Table 3). No statistically significant differences were found in mean daily intakes between the FFQ-1 and the FFQ-2. Although differences in consumption values were slight, offee intake

was lower as reported on the FFQ-B compared with the mean of FFQ-1 and FFQ-2 (mean = 0.69 cups/day vs. 0.77 cups/day, geometric mean = 0.10 vs. 0.11; $P = 0.02$), and tea intake was higher as reported on the FFQ-B (0.47 cups/ day vs. 0.38 cups/day; geometric mean = 0.10 vs. 0.09; $P = 0.04$). Cross-classification between the FFQ-1 and the FFQ-2 showed little severe misclassification (i.e., women assigned to extreme categories), almost perfect agreement for coffee (weighted $\kappa = 0.88$), substantial agreement for caffeine, tea, and soda (weighted $\kappa = 0.76$, 0.62, and 0.72, respectively), and moderate agreement for coffee drinks/ cocoa ($\kappa = 0.56$) (Table 4). Similar levels of agreement were found between the FFQ-B and the mean of FFQ-1 and FFQ-2.

There was no significant variation in caffeine consumption over the 2 menstrual cycles (cycle 1, $P = 0.88$; cycle 2, $P = 0.99$); coffee drinks/cocoa (cycle 1, $P = 0.99$; cycle 2, $P = 0.95$); tea (cycle 1, $P = 0.90$; cycle 2, $P = 0.76$); and soda (cycle 1, $P = 0.71$; cycle 2, $P = 0.87$) (Figure 3). Caffeine consumption reported on the daily diary was also consistent across the menstrual cycle ($P = 0.97$ for both cycles 1 and 2).

DISCUSSION

We showed that although caffeine and caffeinated beverage intakes were highly correlated among measurement tools in the BioCycle Study, absolute intakes differed significantly. Caffeine and coffee intakes were overestimated and soda intake was underestimated in the FFQ compared with the 24HDR, and caffeinated coffee intake was underestimated in the 24HDR compared with the corresponding day's daily diary. Although the FFQ is adequate for ranking women on their caffeine and caffeinated beverage exposures, it may not appropriately classify exposure on the basis of clinically relevant cutpoints. We demonstrated that caffeine intake as reported in the FFQ was consistent over the 2 menstrual cycles under study and showed a consistent pattern of intake over the previous 6 months as reported by the baseline FFQ compared with reports while under observation. Our analysis of the caffeine and caffeinated beverage intakes as reported in the 24HDR and the daily diary further supports our finding that caffeine intake was relatively consistent over the course of the menstrual cycle.

FFQ- and 24HDR-reported caffeine and caffeinated beverage intakes were more highly correlated than in previous validation studies. Prior population-based studies of women demonstrated deattenuated correlations of between 0.64 and 0.76 (13–15, 28). Given that correlations above 0.50 between a dietary instrument (such as the FFQ) and a reference method (such as the 24HDR or the daily diary) indicate that the instrument can reliably rank persons (21), both ours and previous studies support the FFQ as a valid instrument to rank caffeine intake. The FFQ's ability to rank persons for caffeinated beverages is not surprising because participants more easily recall frequently consumed foods and beverages (21).

Although adequate ranking of subjects may be sufficient for many epidemiologic analyses (14, 15), assessments of

Table 3.   Usual Daily Intakes of Caffeine and Caffeinated Beverages Calculated From the FFQ[a] in the BioCycle Study, 2005–2007

	FFQ-1		FFQ-2		P Value[b]	Correlation (R)[c]
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
Caffeine, mg/day	112.6 (130.8)	68.7 (16.4–157.2)	114.5 (136.3)	71.5 (15.6–153.0)	0.92	0.86
Coffee, cups/day[d,e]	0.75 (1.27)	0.06 (0.00–1.00)	0.79 (1.37)	0.06 (0.00–1.00)	0.34	0.94
Coffee drinks/cocoa, cups/day[f]	0.15 (0.32)	0.06 (0.00–0.14)	0.15 (0.39)	0.02 (0.00–0.14)	0.48	0.72
Tea, cups/day	0.39 (0.79)	0.06 (0.00–0.39)	0.37 (0.84)	0.06 (0.00–0.39)	0.10	0.76
Soda, cups/day	0.42 (0.78)	0.12 (0.03–0.39)	0.41 (0.74)	0.12 (0.03–0.39)	0.86	0.84
	FFQ-B		**FFQ-1&2**			
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
Caffeine, mg/day	114.5 (140.4)	71.8 (20.6–150.8)	113.4 (128.6)	70.9 (17.6–152.1)	0.66	0.86
Coffee, cups/day	0.69 (1.21)	0.06 (0.00–1.00)	0.77 (1.28)	0.10 (0.00–1.00)	0.02	0.94
Coffee drinks/cocoa, cups/day	0.19 (0.59)	0.06 (0.00–0.14)	0.15 (0.32)	0.03 (0.00–0.14)	0.15	0.79
Tea, cups/day	0.47 (1.01)	0.09 (0.00–0.39)	0.38 (0.76)	0.09 (0.00–0.39)	0.04	0.76
Soda, cups/day	0.46 (0.82)	0.14 (0.03–0.39)	0.41 (0.69)	0.12 (0.03–0.45)	0.30	0.82

Abbreviations: FFQ-1, food frequency questionnaire captured at the end of menstrual cycle 1; FFQ-2, food-frequency questionnaire captured at the end of menstrual cycle 2; FFQ-B, food frequency questionnaire captured at baseline; FFQ-1&2, mean of FFQ-1 and FFQ-2 over 2 menstrual cycles; IQR, interquartile range; SD, standard deviation.

[a] $n = 224$ for comparison between FFQ-1 and FFQ-2; $n = 222$ for comparison between FFQ-B and the mean of FFQ-1 and FFQ-2.

[b] Wilcoxon's matched-pairs signed-rank test.

[c] Pearson's product-moment correlation coefficient on log-transformed data.

[d] Coffee is all types but "not lattes or mochas."

[e] 1 cup = 237 mL.

[f] Coffee drinks/cocoa includes "latte, cappuccino, mocha, or hot chocolate."

absolute intakes are necessary for formulating recommended levels of consumption and for comparability among studies (15). We found that the mean caffeine intakes reported in the 24HDRs were lower than those reported in the FFQs over the same time period. Our findings agree with those of a previous study of nonpregnant American women, which compared daily coffee intake as reported on the dietary record and the FFQ (1.8 vs. 2.4 cups, respectively) (15). Although we could not compare the coffee intake (caffeinated and decaffeinated) reported in the FFQ with the coffee intake (caffeinated only) reported in the daily diary, given that BioCycle Study participants consumed predominately caffeinated coffee (only 1% of subjects reported consuming exclusively decaffeinated

Table 4.   Cross-classification and κ Coefficient of the FFQ-1 and the FFQ-2 Categories[a] of Daily Caffeine and Caffeinated Beverage Intakes ($n = 224$) in the BioCycle Study, 2005–2007

|	Same Category		Adjacent Category		Extreme Category		Weighted κ	95% CI
	No.	%	No.	%	No.	%		
Caffeine, mg/day	158	71	64	29	0	0	0.76	0.70, 0.81
Coffee, cups/day[b,c]	190	85	14	6	0	0	0.88	0.84, 0.92
Coffee drinks/cocoa, cups/day[d]	144	64	37	17	4	2	0.56	0.47, 0.64
Tea, cups/day	150	67	40	18	10	4	0.62	0.53, 0.70
Soda, cups/day	155	69	60	42	1	0.004	0.72	0.66, 0.79

Abbreviations: CI, confidence interval; FFQ-1, food frequency questionnaire captured at the end of menstrual cycle 1; FFQ-2, food frequency questionnaire captured at the end of menstrual cycle 2.

[a] Caffeine intake is divided into 4 categories on the basis of quartiles; intake of caffeinated beverages is divided into 4 categories: nonconsumers and tertiles.

[b] Coffee is all types but "not lattes or mochas."

[c] 1 cup = 237 mL.

[d] Coffee drinks/cocoa includes "latte, cappuccino, mocha, or hot chocolate."
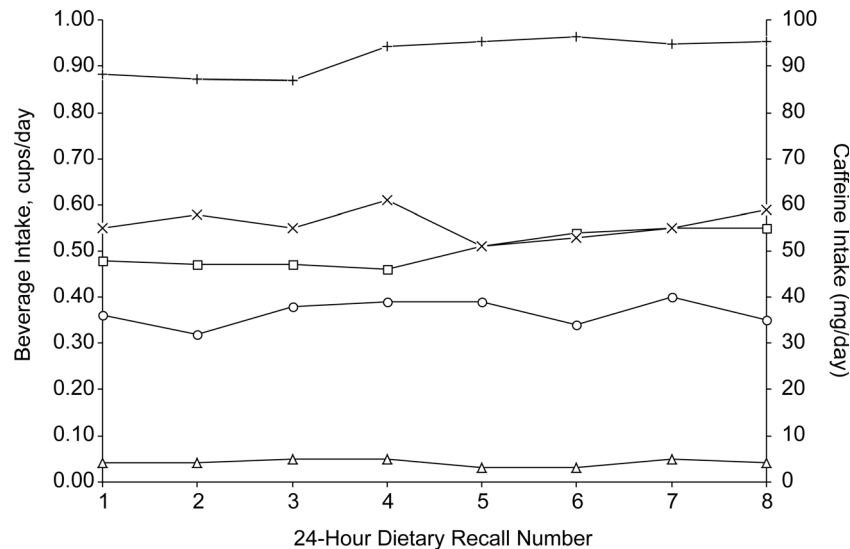
Figure 3. Usual daily intakes of caffeine and caffeinated beverages calculated from eight 24-hour dietary recalls (24HDRs) ($n = 258$) over 2 menstrual cycles in the BioCycle Study, 2005–2007. One cup = 237 mL. Coffee includes all types but "not lattes or mochas." Coffee drinks/cocoa includes "latte, cappuccino, mocha, or hot chocolate." Plus sign, caffeine; square, coffee; triangle, coffee drinks/cocoa; circle, tea; multiplication sign, soda.

coffee), the concordance of the FFQ and the daily diary warrants further research, ideally with a caffeine biomarker, to determine the validity of the FFQ for assessing caffeine exposure among premenopausal women.

One explanation for the difference in the reported intakes between the 24HDR and the FFQ relates to variation in weekday versus weekend consumption. The majority (90%) of the 24HDRs in our study were completed on weekdays and, among this relatively young population, caffeine intake may occur more frequently on weekends because of its correlation with alcohol intake, which occurred more frequently on the weekends in the BioCycle Study. Stratification by weekend versus weekday recall for the 134 participants who had a weekend recall, however, showed no significant differences for caffeine, coffee, coffee drinks/cocoa, tea, or soda (Wilcoxon's matched-pairs signed-ranks test: $P = 0.65, 0.42, 0.40, 0.36$, and 0.60, respectively). The differences in absolute intakes as reported on the FFQ, the 24HDR, and the daily diary may more likely be attributable to the tendency to overreport socially desirable foods and beverages and underreport less healthy foods (15). The standardized, multipass method of a 24HDR may better correct for this bias compared with a self-administered FFQ or the daily diary. Because participants were instructed to report in their daily diaries the total number of cups of caffeinated coffee consumed, rounding up of caffeinated coffee intake may have occurred. Coffee has been publicized to contain antioxidants and to have chemopreventive properties, which could account for the higher reports of coffee intake. Negative publicity about soda may explain our finding of a lower reported consumption in the FFQ compared with the 24HDR.

Classification analyses for caffeine have been conducted in 1 other study by using the FFQ with nearly identical results (weighted $\kappa = 0.64$) despite a lower mean caffeine intake of 114 (standard deviation, 128) mg in our study compared with a mean of 143 (standard deviation, 105) mg in their study (14). We found that the FFQ reliably distinguished extreme caffeine intake, as documented previously (14). No other study has assessed the sensitivity, specificity, and positive and negative predictive values of the FFQ for caffeine intake on the basis of recommended limits of intake. If we assume that the 24HDR accurately assessed caffeine intake, then use of the FFQ as a "screening" tool would wrongly categorize 3% of women as below the recommended range and 8% of women as above the recommended range of intake. If we assume that the daily diary accurately assessed caffeine intake, then use of the 24HDR as a screening tool would wrongly categorize 2% of women as below the recommended range and 27% of women as above the recommended range of intake.

This is the first study to investigate the validity of the FFQ for reporting the intake of coffee drinks/cocoa. Analyses of specific foods or beverages (e.g., coffee drinks) instead of nutrients (e.g., caffeine) are useful for detecting questionnaire weaknesses and for informing potential questionnaire modifications (15). Average intakes of coffee drinks/cocoa as reported for the FFQ and the 24HDR were weakly correlated, indicating that the FFQ poorly measures these beverages, perhaps because multiple beverages (i.e., lattes, cappuccinos, mochas, and hot chocolate) are collapsed into 1 category. If the research aim is to assess caffeine intake, a more detailed caffeine assessment tool, such as the Nutrition Assessment Shared Resource Caffeine Questionnaire (29),

should be considered. Although there is no plan to validate the Nutrition Assessment Shared Resource Caffeine Questionnaire, we believe that a thorough caffeine assessment tool should be validated among premenopausal American women to improve caffeine exposure assessment in this population, as has been done in the United Kingdom (8).

The FFQ, the 24HDR, and the daily diary all showed that caffeine and caffeinated beverage intakes did not vary significantly for BioCycle Study participants, both between the baseline and the study period as well as over the course of 2 menstrual cycles. This indicates that caffeine and caffeinated beverage intakes were habitual among this relatively young premenopausal cohort and were not influenced by study enrollment or menstrual cycle phase. Although caffeine metabolism has been shown to vary across different phases of the menstrual cycle, this does not appear to influence caffeine intake behavior (11).

Although the BioCycle Study had several strengths including high compliance to the study protocol and the ability to evaluate 3 commonly used self-report methods to assess caffeine intake across multiple time points during a relevant window for this population, our study was limited by several factors, including the use of the FFQ-GSEL, which does not distinguish among caffeinated and decaffeinated coffee, teas, and sodas, compared with other FFQs (15, 22). Although this would affect differences in reported caffeine intakes between the FFQ and 24HDR, it would not have affected differences in beverage intake because we combined caffeinated and decaffeinated beverages reported in the 24HDR to allow direct comparison with the FFQ. Additionally, assessing caffeine intake by self-report is difficult because of the heterogeneity of caffeine content in beverages and the inter-variation in caffeine metabolism. Although we show that overall caffeine and caffeinated beverage intakes did not vary, caffeine metabolism may change over the menstrual cycle (11). To improve caffeine exposure assessment among premenopausal women, future studies using a combination of self-reported intake and biomarkers (e.g., caffeine, paraxanthine, theobromine, and theophylline) may increase precision and help to better measure caffeine dose.

In summary, we showed that although the intakes of caffeine and caffeinated beverages reported on the FFQ, the 24HDR, and the daily diary are highly correlated and have acceptable relative limits of agreement, absolute intakes differ significantly among measurement tools. These results highlight the importance of considering potential misclassification of caffeine exposure when assessing its effect on premenopausal women's health. Although we show that caffeinated beverage intake does not vary over the menstrual cycle, we did not assess differences in caffeine metabolism over the menstrual cycle. Further explorations examining the relationship between self-reported measures of caffeine intake and biomarkers of caffeine concentrations are needed.

REFERENCES

1. Smith BD, White T, Shapiro R. The arousal drug of choice: sources and consumption of caffeine. In: Smith BD, Gupta U, Gupta BS, eds. *Caffeine and Activation Theory Effects on Health and Behavior*. Boca Raton, FL: Taylor & Francis Group; 2007:9–40.
2. 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(10):2549–2576.
3. Nkondjock A. Coffee consumption and the risk of cancer: an overview. *Cancer Lett*. 2009;277(2):121–125.
4. Higdon JV, Frei B. Coffee and health: a review of recent human research. *Crit Rev Food Sci Nutr*. 2006;46(2):101–123.
5. Chavarro JE, Rich-Edwards JW, Rosner BA, et al. Caffeinated and alcoholic beverage intake in relation to ovulatory disorder infertility. *Epidemiology*. 2009;20(3):374–381.
6. Bracken M, Triche E, Grosso L, et al. Heterogeneity in assessing self-reports of caffeine exposure: implications for studies of health effects. *Epidemiology*. 2002;13(2):165–171.
7. McCusker RR, Goldberger BA, Cone EJ. Caffeine content of specialty coffees. *J Anal Toxicol*. 2003;27(7):520–522.
8. Boylan SM, Cade JE, Kirk SF, et al. Assessing caffeine exposure in pregnant women. *Br J Nutr*. 2008;100(4):875–882.

9. 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(1):110–113.

10. Kotsopoulos J, Eliassen AH, Missmer SA, et al. Relationship between caffeine intake and plasma sex hormone concentrations in premenopausal and postmenopausal women. *Cancer*. 2009;115(12):2765–2774.

11. Vo HT, Smith BD, Elmi S. Menstrual endocrinology and pathology: caffeine, physiology, and PMS. In: Smith BD, Gupta U, Gupta BS, eds. *Caffeine Activation Theory: Effects on Health and Behavior*. Boca Raton, FL: Taylor & Francis Group; 2007:181–197.

12. 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(1):21–24.

13. London S, Willett W, Longcope C, et al. Alcohol and other dietary factors in relation to serum hormone concentrations in women at climacteric. *Am J Clin Nutr*. 1991;53(1):166–171.

14. Lucero J, Harlow BL, Barbieri RL, et al. Early follicular phase hormone levels in relation to patterns of alcohol, tobacco, and coffee use. *Fertil Steril*. 2001;76(4):723–729.

15. Jain MG, Rohan TE, Soskolne CL, et al. Calibration of the dietary questionnaire for the Canadian Study of Diet, Lifestyle and Health cohort. *Public Health Nutr*. 2003;6(1):79–86.

16. Bolca S, Huybrechts I, Verschraegen M, et al. Validity and reproducibility of a self-administered semi-quantitative food-frequency questionnaire for estimating usual daily fat, fibre, alcohol, caffeine and theobromine intakes among Belgian post-menopausal women. *Int J Environ Res Public Health*. 2009;6(1):121–150.

17. Salvini S, Hunter DJ, Sampson L, et al. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. *Int J Epidemiol*. 1989;18(4):858–867.

18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Int J Nurs Stud*. 2010;47(8):931–936.

19. Wactawski-Wende J, Schisterman EF, Hovey KM, et al. BioCycle study: design of the longitudinal study of the oxidative stress and hormone variation during the menstrual cycle. *Paediatr Perinat Epidemiol*. 2009;23(2):171–184.

20. Fred Hutchinson Cancer Research Center. Nutrition assessment shared resources: food frequency questionnaires. Seattle, WA: Fred Hutchinson Cancer Research Center; 2012. http://sharedresources.fhcrc.org/services/food-frequency-questionnaires-ffq). (Accessed May 3, 2012).

21. 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(suppl 4):1130S–1137S.

22. Beaton GH, Milner J, Corey P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am J Clin Nutr*. 1979;32(12): 2546–2559.

23. Euser AM, Dekker FW, le Cessie S. A practical approach to Bland-Altman plots and variation coefficients for log transformed variables. *J Clin Epidemiol*. 2008;61(10): 978–982.

24. Willett W. *Nutritional Epidemiology*. 2nd ed. New York, NY: Oxford University Press; 1998.

25. Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*. 1985;122(1): 51–65.

26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1): 159–174.

27. Carl J, Hill DA. Preconception counseling: make it part of the annual exam. *J Fam Pract*. 2009;58(6):307–314.

28. Addicott MA, Yang LL, Peiffer AM, et al. Methodological considerations for the quantification of self-reported caffeine use. *Psychopharmacology (Berl)*. 2009; 203(3):571–578.

29. Fred Hutchinson Cancer Research Center. Nutrition assessment shared resources: specific food questionnaires: caffeine questionnaire (supplemental beverages). Seattle, WA: Fred Hutchinson Cancer Research Center. (http://sharedresources.fhcrc.org/documents/caffeine-questionnaire). (Accessed May 3, 2012).