Published in final edited form as:

Paediatr Perinat Epidemiol. 2011 September ; 25(5): 448–459. doi:10.1111/j.1365-3016.2011.01204.x.

Realignment and multiple imputation of longitudinal data: an application to menstrual cycle data

Sunni L. Mumford^a, Enrique F. Schisterman^a, Audrey J. Gaskins^a, Anna Z. Pollack^a, Neil J. Perkins^a, Brian W. Whitcomb^b, Aijun Ye^a, and Jean Wactawski-Wende^c

^aDivision of Epidemiology, Statistics, and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Rockville, MD

^bDivision of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts at Amherst, Amherst, MA

^cDepartment of Social and Preventive Medicine, University at Buffalo, State University of New York, Buffalo, NY, USA

Summary

Reproductive hormone levels are highly variable among premenopausal women during the menstrual cycle. Accurate timing of hormone measurement is essential, especially when investigating day- or phase-specific effects. The BioCycle Study used daily urine home fertility monitors to help detect the luteinising hormone (LH) surge in order to schedule visits with biologically relevant windows of hormonal variability. However, as the LH surge is brief and cycles vary in length, relevant hormonal changes may not align with scheduled visits even when fertility monitors are used. Using monitor data, measurements were reclassified according to biological phase of the menstrual cycle to more accurate cycle phase categories. Longitudinal multiple imputation methods were applied after reclassification if no visit occurred during a given menstrual cycle phase. Reclassified cycles had more clearly defined hormonal profiles, with higher mean peak hormones (up to 141%) and reduced variability (up to 71%). We demonstrate the importance of realigning visits to biologically relevant windows when assessing phase- or day-specific effects and the feasibility of applying longitudinal multiple imputation methods. Our method has applications in settings where missing data may occur over time, where daily blood sampling for hormonal measurements is not feasible, and in other areas where timing is essential.

Keywords

menstrual cycle; multiple imputation; statistical methodology; oestradiol 2; luteinising hormone; diet; fibre; missing data

Introduction

Reproductive hormones are highly variable among premenopausal women, reflecting the dynamic biological processes underlying the menstrual cycle. Timing sample collection to biologically relevant windows is critical for estimating day- or phase-specific effects in premenopausal women where hormone levels vary across cycle phases, and exposures may influence these processes differently across phases of the cycle. Identifying and capturing these biologically relevant windows represents a major challenge in reproductive epidemiology.¹

To date, researchers have used a variety of sampling approaches in an attempt to time measurement of hormone levels. The gold standard approach involves daily collection and storage of first-morning urine specimens; however, in large cohorts, daily sample collection entails substantial cost and high participant burden, which can influence participation and lead to missing data because of non-compliance and dropout. As an alternative approach for epidemiological studies, biospecimen collection and storage occurs at a handful of carefully timed clinic visits scheduled to occur at key times of hormonal variability, minimising participant burden. One of the most common approaches is to schedule biospecimen collection based on a standard 28-day cycle.¹ However, because of the considerable variability in menstrual cycle length (even within a woman), the use of a 28-day calendar to time visits often leads to misclassification of cycle phases.^{2,3}

Recently, studies have attempted to better standardise visits across the menstrual cycle by using predictors of ovulation,⁴⁻⁷ a transition that divides the menstrual cycle into follicular and luteal phases.⁸ Fertility monitors provide a non-invasive approach to predict ovulation with minimal participant burden using daily morning urine.³ Monitors measure oestrone-3-glucuronide and luteinising hormone (LH) in urine and detect rises in oestrone-3-glucuronide and the LH surge, which occurs approximately 24–36 h prior to ovulation.⁸ When ovulation is correctly identified, measurements taken before and after ovulation can be correctly classified as occurring during the follicular or luteal phases, respectively. However, the LH surge that precedes ovulation is short-lived and therefore not easily detected if test days are skipped. When timing of ovulation is not correctly identified, mid-cycle visits may not be scheduled at the appropriate time, leading to misclassification of the cycle phase for some visits. Even in cases where the LH surge is identified, if women do not have a clinic visit that same day or the following day, data could be influenced by misclassification of the cycle phase. Thus, although fertility monitors greatly improve cycle phase classification, misclassification may occur.

Information on hormone levels may be used to correct for this misclassification by *realigning* the data. Realignment of clinic visits ensures that hormonal measurements are classified to the correct menstrual cycle phase and is critical when researchers intend to make phase-specific or acute comparisons of hormones. A complication of this approach for realignment of cycle visits is the missing data that occurs when data are reclassified to the correct cycle phase. In theory, multiple imputation methods may be used to estimate hormone levels on missing cycle visits to alleviate missing data problems that result from realignment; however, this has yet to be demonstrated in practice.

In this paper, we consider realignment of longitudinal hormonal data to the correct cycle phase in the BioCycle Study. We describe use of data from fertility monitors and serum hormone levels at each visit to inform realignment, so clinic visit data are correctly classified by menstrual cycle phase. We also evaluate the effectiveness of multiple imputation to account for missing data generated by the realignment process. Lastly, we demonstrate the impact of realigning menstrual cycle visits on estimation of day- and phase-specific effects in the context of the association between daily fibre intake and reproductive hormone levels among premenopausal women.

Methods

The BioCycle Study

The BioCycle Study was a prospective cohort study which recruited 259 healthy, premenopausal female volunteers, aged 18 to 44 years from western New York between 2005–07, and followed them for one ($n = 9$) or two ($n = 250$) cycles.⁹ Inclusion and exclusion criteria for BioCycle are described elsewhere in detail.¹⁰ Briefly, women with a self-reported cycle length between 21 and 35 days during the past 6 months and without conditions that might affect menstrual cycle function were eligible. Study participants were highly compliant; the number of visits ranged from five to eight per cycle, with more than 93% completing seven or eight visits per cycles. The University at Buffalo Health Sciences Institutional Review Board (IRB) approved the study and served as the IRB designated by the National Institutes of Health for this study under a reliance agreement. All participants provided written informed consent.

Fertility monitors

Study visits were originally scheduled to occur during specific phases of the menstrual cycle using an algorithm accounting for each woman's self-reported cycle length (Table 1), with the day of the mid-cycle visits adjusted based on fertility monitor data. Visits were scheduled to correspond to menstruation (visit 1), the mid follicular phase (visit 2), three visits during the periovulatory phase (visits 3, 4 and 5, with visit 4 corresponding to the LH surge), and the early (visit 6), mid (visit 7) and late luteal phases (visit 8). Fertility monitors (Clearblue Easy™ Fertility Monitor, Inverness Medical, Waltham, MA, USA) measured oestrone-3-glucuronide and LH in urine, starting on calendar day 6 after menses and continuing for 10–20 days, depending on whether the woman reached peak levels on the monitor.³ The monitor determines 'peak fertility' on the basis of predetermined cut-points during the first cycle, but in subsequent cycles adjusts the cut-point criteria according to the woman's specific hormone levels. Monitor indications of low, high and peak fertility were used to time mid-cycle visits. If the monitor indicated 'peak fertility' on a day without a scheduled visit, the women were asked to come in that morning and the following two mornings. The remaining visits were scheduled based on the algorithm described in Table 1 adjusting for self-reported cycle length. Monitors were adapted to capture data on a computer chip which was downloaded for use in this study. Eighty-four per cent of women had complete fertility monitor data for both cycles, and among the ovulatory women in the study who reached 'peak fertility' on the monitor, we observed that serum and urine LH peaks occurred within one calendar day in 81% of cycles.

Hormone assessment

Oestradiol, progesterone, LH and follicle-stimulating hormone (FSH) were measured in fasting serum samples collected at each clinic visit at Kaleida Health Laboratories (Buffalo, NY, USA). Oestradiol was measured by radioimmunoassay, and progesterone, LH and FSH were measured using solid phase competitive chemiluminescent enzymatic immunoassay by Specialty Laboratories, Inc. (Valencia, CA, USA) on the DPC Immulite®2000 analyser (Siemens Medical Solutions Diagnostics, Deerfield, IL, USA). Across the study period, the coefficient of variation for these tests reported by the laboratory were <10% for oestradiol, <5% for LH and FSH, and <14% for progesterone. Anovulatory cycles (defined as cycles with progesterone levels ≤ 5 ng/mL, and no observed serum LH peak during the mid- or late luteal phase visits, $n = 42$) were excluded from this analysis, as their hormonal profiles are distinctly different and timing of data collection does not play a role as there is no hormonal variability.¹¹

Assessment of phase misclassification

Despite the use of fertility monitors to time mid-cycle visits and the high compliance with the study protocol, *post hoc* evaluation of LH levels revealed peak levels in samples collected at each of the eight visits for one or more of the cycles, suggesting that some visits were not correctly timed according to cycle phase. The LH surge was correctly captured on cycle visit 4 (corresponding to the LH surge during the periovulatory period), in 27.7% of visits in cycle 1, and 32.8% of visits in cycle 2 (Table 2). The majority of misclassified peak LH values occurred between visits 3 and 5, 78% and 83% for cycles 1 and 2, respectively. Misclassification could be due to several factors. In some cases, women had low peak serum LH values, despite reporting to the clinic on the assigned day based on her monitor indications. Alternatively, in some women the LH surge was not captured on the monitor because of missed tests or long cycle length and thus clinic visits were not timed appropriately. It is unlikely that in these cases the serum LH measurement was picking up hCG (human chorionic gonadotropin), as each of the women had a negative pregnancy test at baseline and at the end of each cycle. Lastly, upon review of both monitor and serum data, there was no evidence that women had multiple LH surges throughout the cycle.

Realignment algorithm

To ensure that women were compared during the same cycle phases (e.g. standardised to the same biological time scale) and to decrease misclassification, we realigned cycle visits to the appropriate menstrual cycle phase based on urinary data from fertility monitors and serum hormone levels from clinic visits. Visits were realigned to occur during eight specific cycle phases: menstruation (M), mid follicular (F1), periovulatory (O1, O2, O3), and early (L1), mid (L2) and late luteal phase (L3). Table 3 summarises the algorithm used to realign menstrual cycle visit data, and Figure 1 provides a graphical description of the algorithm for realigning cycles with an observed LH peak on visit 6. For correctly timed visits (i.e. where the LH peak occurred on visit 4, corresponding to the LH surge on the monitor and phase O2), values were not changed. For mistimed visits, the algorithm reclassified the visit with the LH peak to phase O2, and reclassified the remaining visits accordingly. For example, after a review of the data, if the LH peak was observed on visit 3 rather than visit 4, then

data for visit 3 were manually reclassified as visit 4 (phase O2), and data for visit 3 were considered to be missing. No changes were made to the luteal phase visits (i.e. visits 5–8) when the LH peak was misclassified by only one clinic visit. To reduce variability, data from adjacent visits were averaged to estimate hormone levels on certain visits when the amount of time between visits was minimal. Among cycles with late LH peaks (visits 6–8), visits were realigned in a similar manner by reclassifying the LH peak to phase O2, and then reclassifying the remaining visits as outlined in Table 3 and displayed in Figure 1.

The algorithm described is inadequate for classifying cycle phases when the LH peak was not captured in serum. As such, all cycles with an observed peak LH level <20 ng/mL or a peak LH in the early follicular phase (visit 1 or 2) were evaluated on a case-by-case basis. These clinic visits were realigned based on urinary data from the fertility monitor, calendar dates of cycle visits, serum hormone levels at each visit and menses information (see Figure 2 for details).

Multiple imputation

Although there was excellent compliance with the study protocol, realignment induced phase-specific missing data on multiple clinic visits for a large percentage of the women in this study. Multiple imputation was used to address the missing data problems generated from reclassifying cycle visits, as they provide a useful strategy for dealing with incomplete multivariate data.^{12,13} Briefly, multiple imputation makes use of available distributional and covariate information to inform imputation for each missing value. Multiple datasets are created, as each missing value is replaced with a set of random plausible values conditional on known covariates and known distributional information. Analysis is completed on each imputed dataset and then results are combined taking into account the uncertainty of the imputation to ensure valid statistical inference. Techniques for multiple imputation have been described in detail elsewhere,^{12–16} and have been successfully applied to menstrual cycle data.¹⁷ These procedures assume that data are missing at random (MAR) or that missing data depend on observed covariates but not on the unobserved data. Although untestable, the MAR assumption is reasonable in the current setting, as information on hormone levels is available from visits before and after the missing values, as well as information from other hormones which are highly correlated and biologically informative because of the complex feedback mechanisms which regulate the menstrual cycle.⁸ The multivariate imputation by chained equations (MICE) package for R was applied to create 10 datasets, imputing values for oestradiol, progesterone, LH and FSH missing due to the realignment algorithm. Data were assumed to be MAR after conditioning upon age, race, body mass index and adjacent hormone measurements.

Application: fibre and reproductive hormones

We will demonstrate an application of the realignment algorithm and multiple imputation to evaluate relationships between fibre intake and hormone levels in the BioCycle Study. High-fibre diets are recommended because of their association with reduced risk of chronic disease.^{18,19} However, research of the effect of fibre intake on endogenous reproductive hormone levels in younger women is limited, even though it is of particular interest given the effect of these hormones on conception and pregnancy maintenance.^{20,21} Here we

extend a previous analysis of the effect of fibre on hormone levels¹¹ by using the realigned data and multiple imputation to evaluate the impact of realignment on estimation of day- and phase-specific effects of fibre intake on hormone levels.

Statistical analysis

To demonstrate the impact of the realignment, the mean and variance of hormone levels at each clinic visit were compared before realignment, after realignment without multiple imputation, and after realignment with multiple imputation. Hormones were log transformed for normality and geometric means are presented for comparison. Dietary intake was assessed four times per cycle using a 24-h dietary recall and averaged across each cycle (Nutrition Data System for Research software version 2005, Nutrition Coordinating Center, University of Minnesota, Minneapolis, MN, USA). Linear mixed effects models with random intercepts on the log scale of the hormones were used to evaluate the association between hormone concentrations and average fibre intake per cycle (in 5 g/day increments). Models were adjusted for age, race, energy intake and vitamin E, for consistency with previous work.¹¹ We compared estimation of day- and phase-specific effects of fibre intake on hormone levels both before and after realignment and imputation. Specifically, we compared the effect of fibre intake on LH and FSH levels during the LH surge (phase O2) and around ovulation (phases O1, O2, O3), and the effect on progesterone throughout the cycle vs. the luteal phase (phases L1, L2, L3). All statistical analyses took multiple cycles per woman and multiple imputation into account and were carried out using SAS version 9.1 (SAS Institute, Cary, NC, USA).

Results

Realignment algorithm

Of 509 total cycles observed in the BioCycle Study, 467 were ovulatory and were considered in our analyses. Of the ovulatory cycles, 328 (70%) were realigned so that the peak LH occurred on the LH surge visit (phase O2). Of those that were realigned, 267 (81.4%) were realigned according to the algorithm in Table 3 and 61 (18.6%) were realigned on a case-by-case basis. Most of the realignment required only minor adjustments as the LH peak occurred on visit 3 or 5 (42%). Under the original data structure, data were missing for 173 hormone measurements from 141 cycles (4072 total visits). Under the realigned data structure, data were missing for 735 hormone measurements from 328 cycles (thus, 562 missing hormone measurements were due to realignment).

Realigned cycles had higher mean peak hormone values and reduced visit-specific variability (Table 4) throughout the cycle, as compared with the original data. Realignment resulted in a more clearly defined hormonal profile (Figure 3). The largest geometric mean changes were observed on visit 4, intended to capture the LH surge (phase O2); specifically, mean LH increased by 141%, oestradiol increased by 50% and FSH increased by 52%. Luteal phase progesterone (visit 7) increased by 34%. Greater change in mean peak values occurred in cycle 1 than 2 (data not shown). The variance of log hormone levels was reduced after realignment at each clinic visit. After realignment, variances of LH, FSH, oestradiol and progesterone decreased by up to 62%, 48%, 43% and 71%, respectively.

Results were similar after multiple imputation, with the additional impact of imputation being small in relation to the impact after realignment.

Application: fibre and reproductive hormones

We observed significant inverse ovulation phase-specific associations (phases O1, O2, O3) between fibre intake and LH levels both before and after realignment, although the realigned and imputed data showed improved precision (Table 5). When we evaluated the association on the day of the LH surge (phase O2), we observed even stronger associations with fibre intake, in particular using the realigned and imputed data. We observed significant day- and phase-specific associations between fibre intake and FSH only when using the realigned and imputed data, which is most likely to be a consequence of improved precision following realignment. Further, luteal phase-specific associations between fibre intake and progesterone were stronger than associations across the cycle, with smaller standard errors observed after realigning the cycle visits and applying multiple imputation. Associations between fibre and oestrogen across the cycle were similar using each of the three models.

Discussion

Timing sample collection to biologically relevant windows is critical for estimating day- or phase-specific effects, and represents a major challenge in epidemiology. We demonstrated improved cycle phase classification by realigning longitudinal data from a menstrual cycle study, with multiple imputation to account for missing data generated by the realignment process. Reclassified cycles had more clearly defined hormonal profiles, with higher mean peak reproductive hormone levels and reduced variability throughout the cycle, leading to more precise estimation of day- and phase-specific effects. We applied our methodology to evaluate the association between fibre intake and reproductive hormone levels as previous research has been inconsistent. While an association between oestradiol levels and fibre intake is established,^{11,22–27} the association between fibre intake and LH levels has only been evaluated in three previous studies,^{11,22,28} with only one study observing a significant inverse association.¹¹ Issues of timing of hormone measurement to appropriate menstrual cycle phase may have influenced the results of these studies, highlighting the importance of timing when assessing phase-specific effects.

To capture biologically relevant windows when studying factors affecting menstrual cycle function, techniques to standardise sample collection should utilise ovulation predictors. Nevertheless, studies with well-timed measurements and high compliance remain subject to phase misclassification, and even studies with daily biospecimen collection are subject to missing data. The strategies we describe include realignment for phase classification and estimation of day- and phase-specific effects and multiple imputation to address missing data problems. Multiple imputation provides an alternative to procedures that require complete case analysis and has the advantage over simple imputation in that it takes the uncertainty of the imputation into account, thus avoiding variance inflation and invalid inferences. Hormone levels are strongly associated with age, race, body mass index and hormone levels from previous and subsequent clinic visits, which make multiple imputation particularly well suited in this context.^{1,8}

Multiple imputation is limited to situations where missing data can be assumed to be MAR. Missing data generated by the realignment process can be assumed to be MAR as realignment was based on availability of fertility monitor data to time visits, and hormone data from adjacent visits were used to inform the imputation. Missing test days on the monitor can be considered to occur at random; it is unlikely that a woman's ovulation status would influence her likelihood of testing. In the current setting, we also assume that the phase relationships are similar between women with predictable cycles and women with more erratic cycles. However, in cases where the missing data generated from realignment are not MAR, information on additional covariates may be needed to meet the MAR assumption. If additional information is not available, other techniques will be necessary. Furthermore, the proposed realignment method is best suited to settings where more frequent measurements or multiple sources of data are available to inform realignment.

The proposed realignment and multiple imputation approach has potential applications in other settings where phase-specific effects are of interest. For example, one might apply these techniques to the study of fetal growth, adolescent development, infectious disease, progression of HIV or other scenarios where exposure effects may vary by phase and timing is important. These are settings where timing of measurement may be based on set calendar times out of necessity or convenience, rather than times of relevant biological changes, which could potentially affect estimation of phase-specific effects. Generally, application could extend to any other settings where there is knowledge of the underlying biological process to inform realignment and sufficient data to properly inform imputation.

In conclusion, our results demonstrate the importance of timing biospecimen collection to biologically relevant windows when assessing day- and phase-specific effects. Even in studies that utilise fertility monitors to time sample collection, misclassification of cycle phase is possible. Realignment of longitudinal menstrual cycle data improves phase classification, and multiple imputation can account for missing data generated by the realignment process. Realigning menstrual cycle data may allow researchers to observe more precise day- and phase-specific effects because of the decrease in variability and misclassification. Our realignment method has applications beyond the scope of menstrual cycle function and may be used in other research settings where timing is essential.

Acknowledgments

We are indebted to all the investigators and staff at the Epidemiology Branch, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development and the University at Buffalo for their respective roles in the study. We would also like to recognise the Bio-Cycle participants for their extraordinary commitment to the study.

Funding

This study was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health.

References

1. Harlow SD, Ephross SA. Epidemiology of menstruation and its relevance to women's health. *Epidemiologic Reviews*. 1995; 17 :265–286. [PubMed: 8654511]

2. Fehring RJ, Schneider M, Raviele K. Variability in the phases of the menstrual cycle. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*. 2006; 35:376–384.
3. 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. *American Journal of Epidemiology*. 2009; 169:105–112. [PubMed: 18974081]
4. Bedford JL, Prior JC, Hitchcock CL, Barr SI. Detecting evidence of luteal activity by least-squares quantitative basal temperature analysis against urinary progesterone metabolites and the effect of wake-time variability. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*. 2009; 146:76–80.
5. Behre HM, Kuhlage J, Gassner C, Sonntag B, Schem C, Schneider HP, et al. Prediction of ovulation by urinary hormone measurements with the home use ClearPlan Fertility Monitor: comparison with transvaginal ultrasound scans and serum hormone measurements. *Human Reproduction*. 2000; 15:2478–2482. [PubMed: 11098014]
6. Rosenberg CR, Pasternack BS, Shore RE, Koenig KL, Toniolo PG. Premenopausal estradiol levels and the risk of breast cancer: a new method of controlling for day of the menstrual cycle. *American Journal of Epidemiology*. 1994; 140:518–525. [PubMed: 8067345]
7. Stanford JB, Smith KR, Dunson DB. Vulvar mucus observations and the probability of pregnancy. *Obstetrics and Gynecology*. 2003; 101:1285–1293. [PubMed: 12798538]
8. Strauss, JF.; Barbieri, RL., editors. *Yen and Jaffe's Reproductive Endocrinology: Expert Consult*. 6. Philadelphia, PA: Saunders, Elsevier; 2009.
9. Schisterman EF, Gaskins AJ, Mumford SL, Browne RW, Yeung E, Trevisan M, et al. Influence of endogenous reproductive hormones on F2-isoprostane levels in premenopausal women: the BioCycle Study. *American Journal of Epidemiology*. 2010; 172:430–439. [PubMed: 20679069]
10. Wactawski-Wende J, Schisterman EF, Hovey KM, Howards PP, Browne RW, Hediger M, et al. BioCycle study: design of the longitudinal study of the oxidative stress and hormone variation during the menstrual cycle. *Paediatric and Perinatal Epidemiology*. 2009; 23:171–184. [PubMed: 19159403]
11. Gaskins AJ, Mumford SL, Zhang C, Wactawski-Wende J, Hovey KM, Whitcomb BW, et al. Effect of daily fiber intake on reproductive function: the BioCycle Study. *American Journal of Clinical Nutrition*. 2009; 90:1061–1069. [PubMed: 19692496]
12. Little, RJ.; Rubin, D. *Statistical Analysis with Missing Data*. 2. New York: Wiley-Interscience; 2002.
13. Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons; 1987.
14. Carlin JB, Galati JC, Royston P. A new framework for managing and analyzing multiply imputed data in Stata. *Stata Journal*. 2008; 8:49–67.
15. Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behavioral Research*. 1998; 33:545–571.
16. van, Buuren S.; Oudshoorn, K. *MICE V1.0 User's manual*. TNO Prevention and Health; 2000. *Multivariate Imputation by Chained Equations*. <http://web.inter.nl.net/users/S.van.Buuren/mi/docs/Manual.pdf> (07.10.2005) [last accessed 18 February 2011]
17. Little RJ, Yosef M, Cain KC, Nan B, Harlow SD. A hot-deck multiple imputation procedure for gaps in longitudinal data on recurrent events. *Statistics in Medicine*. 2008; 27:103–120. [PubMed: 17592832]
18. Institute of Medicine. *Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids*. Washington, D.C: The National Academies Press; 2002.
19. US Department of Agriculture and US Department of Health and Human Services. *Dietary Guidelines for Americans*, 2010. 7. Washington, D.C: US Government Printing Office; 2010.
20. Bussen S, Sutterlin M, Steck T. Endocrine abnormalities during the follicular phase in women with recurrent spontaneous abortion. *Human Reproduction*. 1999; 14:18–20. [PubMed: 10374087]
21. Mason P, Adams J, Morris DV, Tucker M, Price J, Voulgaris Z, et al. Induction of ovulation with pulsatile luteinising hormone releasing hormone. *British Medical Journal*. 1984; 288:181–185. [PubMed: 6419847]

22. Bagga D, Ashley JM, Geffrey SP, Wang HJ, Barnard RJ, Korenman S, et al. Effects of a very low fat, high fiber diet on serum hormones and menstrual function. Implications for breast cancer prevention. *Cancer*. 1995; 76:2491–2496. [PubMed: 8625075]
23. Kaneda N, Nagata C, Kabuto M, Shimizu H. Fat and fiber intakes in relation to serum estrogen concentration in premenopausal Japanese women. *Nutrition and Cancer*. 1997; 27 :279–283. [PubMed: 9101558]
24. Rose DP, Goldman M, Connolly JM, Strong LE. High-fiber diet reduces serum estrogen concentrations in premenopausal women. *American Journal of Clinical Nutrition*. 1991; 54:520–525. [PubMed: 1652197]
25. Woods MN, Barnett JB, Spiegelman D, Trail N, Hertzmark E, Longcope C, et al. Hormone levels during dietary changes in premenopausal African-American women. *Journal of the National Cancer Institute*. 1996; 88:1369–1374. [PubMed: 8827014]
26. Goldin BR, Woods MN, Spiegelman DL, Longcope C, Morrill-LaBrode A, Dwyer JT, et al. The effect of dietary fat and fiber on serum estrogen concentrations in premenopausal women under controlled dietary conditions. *Cancer*. 1994; 74:1125–1131. [PubMed: 8039147]
27. Gann PH, Chatterton RT, Gapstur SM, Liu K, Garside D, Giovanazzi S, et al. The effects of a low-fat/high-fiber diet on sex hormone levels and menstrual cycling in premenopausal women: a 12-month randomized trial (the diet and hormone study). *Cancer*. 2003; 98:1870–1879. [PubMed: 14584069]
28. Chearskul S, Supingklud N, Nitithamyong A, Sirichakwal P. Assessment of hormonal and metabolic effects of dietary fiber in young Thai women. *Journal of the Medical Association of Thailand*. 2006; 89:997–1003. [PubMed: 16881433]

Figure 1.

Example of algorithm for realignment of menstrual cycle visits when the observed luteinising hormone (LH) peak occurs on visit 6; 'x' represents missing data. Visits 1 through 8 correspond to the original data collected. Data were realigned based on the algorithm in Table 3 to correspond to specific menstrual cycle phases: menstruation (M), mid follicular (F1), LH surge (O2), early luteal (L1) and mid luteal (L2). If no data were available during a cycle phase, then the value was set to missing ('x').

Figure 2.

Example of steps used to individually realign cycles with low luteinising hormone (LH) peak levels ($n = 61$ cycles realigned individually).

Figure 3.

Hormonal profiles across the menstrual cycle before and after realignment among women in the BioCycle Study ($n = 259$), Buffalo, NY, USA, 2005–07. Data were realigned to correspond to specific menstrual cycle phases: menstruation (M), mid follicular (F1), periovulatory (O1, O2, O3), and early luteal (L1), mid luteal (L2) and late luteal (L3).

Table 1

Algorithm to schedule clinic visits (in days) based on average self-reported cycle length in the BioCycle Study, Buffalo, NY, USA, 2005–07

Average cycle length (days)	Visit 1 Menses	Visit 2 Mid follicular	Visit 3 Perioviulatory 1	Visit 4 Perioviulatory 2 ^a	Visit 5 Perioviulatory 3	Visit 6 Early luteal	Visit 7 Mid luteal	Visit 8 Late luteal
21	2	4	5	6	7	11	15	20
22	2	4	6	7	8	12	16	21
23	2	5	7	8	9	13	17	22
24	2	5	8	9	10	14	18	23
25	2	6	9	10	11	15	19	24
26	2	6	10	11	12	16	20	25
27	2	7	11	12	13	17	21	26
28	2	7	12	13	14	18	22	27
29	2	7	13	14	15	19	23	28
30	2	8	14	15	16	20	24	29
31	2	8	15	16	17	21	25	30
32	2	9	16	17	18	22	26	31
33	2	9	17	18	19	23	27	32
34	2	10	18	19	20	24	28	33
35	2	10	19	20	21	25	29	34

^a Luteinising hormone surge.

Table 2

Day of maximum serum luteinising hormone (LH) level by cycle visit in the BioCycle Study excluding 42 anovulatory cycles, Buffalo, NY, USA, 2005–07

Visit	Anticipated cycle phase	Cycle 1		Cycle 2	
		n	%	n	%
1	Menses	1	0.4	0	0.0
2	Mid follicular	3	1.3	5	2.2
3	Periovulatory 1	71	30.2	65	28.0
4	Periovulatory 2 (LH surge)	65	27.7	76	32.8
5	Periovulatory 3	47	20.0	52	22.4
6	Early luteal	21	8.9	19	8.2
7	Mid luteal	21	8.9	11	4.7
8	Late luteal	6	2.6	4	1.7
Total		235	100.0	232	100.0

n, number of cycles.

Table 3

Algorithm for realigning menstrual cycle visits to align the day of the luteinising hormone (LH) surge (visit 4) on the standardised LH surge visit (O2) among the 509 observed menstrual cycles of participants ($n = 259$) in the BioCycle Study, Buffalo, NY, USA, 2005–07^a

Standardised cycle phase ^c								
LH peak occurred on visit (<i>n</i>)	Menses (M)	Mid follicular (F1)	Periovulatory 1 (O1)	Periovulatory 2 (LH surge; O2)	Periovulatory 3 (O3)	Early luteal (L1)	Mid luteal (L2)	Late luteal (L3)
Early LH peak	Visit 1 Visit 3 (115) <i>Periovulatory 1</i>	Visit 2	<i>Missing^b</i>	Visit 3	<i>Average of visits 4 & 5</i>	Visit 6	Visit 7	Visit 8
Correctly timed	Visit 1 Visit 4 (139) <i>Periovulatory 2 (LH surge)</i>	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8
Late LH peak	Visit 1 Visit 5 (98) <i>Periovulatory 3</i>	Visit 2	<i>Average of visits 3 & 4</i>	Visit 5	<i>Missing</i>	Visit 6	Visit 7	Visit 8
	Visit 1 Visit 6 (30) <i>Early luteal phase</i>	<i>Average of visits 3, 4 & 5</i>	<i>Missing</i>	Visit 6	<i>Missing</i>	Visit 7	Visit 8	<i>Missing</i>
	Visit 1 Visit 7 (19) <i>Mid luteal phase</i>	<i>Average of visits 3, 4 & 5</i>	<i>Missing</i>	Visit 7	<i>Missing</i>	Visit 8	<i>Missing</i>	<i>Missing</i>
	Visit 1 Visit 8 (5) <i>Late luteal phase</i>	<i>Average of visits 3, 4 & 5</i>	<i>Missing</i>	Visit 8	<i>Missing</i>	<i>Missing</i>	<i>Missing</i>	<i>Missing</i>

n , number of cycles.

^a 61 cycles which lacked a serum LH surge (due to mistimed visits) were shifted individually and 42 anovulatory cycles were not shifted.

^b 'Missing' indicates that after realignment there was no serum collection (visit) during that phase of the cycle such that the reclassified visit was set to missing.

^c If we assume a standard 28-day cycle, the standardised cycle phases would correspond to approximately days 2, 7, 12, 13, 14, 18, 22 and 27, respectively. Alternatively, these visits could be referenced from the day of the LH surge of a 28-day cycle as: onset of menses until day -8 relative to the LH surge, day -7 to -2, day 0, day +1, day +2 to day +7, day +8 to day +11, day +12 to end of cycle.

Comparison of geometric mean (variance) of serum hormone concentrations before and after application of the realignment algorithm in the BioCycle Study, Buffalo, NY, USA, 2005–07

Table 4

Visit	Sequence	n	Oestradiol ^a	Progesterone	LH	FSH
1	Original	507	33.21 (0.23)	0.52 (0.39)	3.86 (0.31)	6.22 (0.20)
	Realigned	506	33.06 (0.22)	0.51 (0.39)	3.87 (0.31)	6.23 (0.20)
	Realigned with MI	5090	33.08 (0.22)	0.51 (0.39)	3.87 (0.31)	6.23 (0.20)
2	Original	507	51.22 (0.36)	0.37 (0.35)	4.69 (0.24)	6.59 (0.12)
	Realigned	494	53.94 (0.31)	0.37 (0.32)	4.91 (0.22)	6.48 (0.11)
	Realigned with MI	5090	54.21 (0.31)	0.37 (0.31)	4.90 (0.22)	6.51 (0.12)
3	Original	509	117.34 (0.60)	0.60 (0.70)	9.36 (0.69)	6.55 (0.25)
	Realigned	298	148.15 (0.46)	0.43 (0.29)	7.64 (0.26)	5.11 (0.13)
	Realigned with MI	5090	165.45 (0.43)	0.42 (0.28)	7.54 (0.25)	4.87 (0.13)
4	Original	509	119.66 (0.54)	0.98 (0.96)	11.59 (0.71)	7.09 (0.30)
	Realigned	443	180.01 (0.41)	0.97 (0.28)	27.94 (0.44)	10.75 (0.24)
	Realigned with MI	5090	182.30 (0.40)	0.98 (0.27)	28.41 (0.42)	10.82 (0.23)
5	Original	503	99.87 (0.46)	1.75 (1.18)	9.56 (0.74)	6.56 (0.32)
	Realigned	340	78.31 (0.26)	2.04 (0.56)	8.37 (0.35)	6.48 (0.22)
	Realigned with MI	5090	78.02 (0.25)	1.84 (0.62)	8.68 (0.34)	6.68 (0.23)
6	Original	498	112.13 (0.31)	4.28 (1.60)	6.59 (0.69)	4.55 (0.34)
	Realigned	488	105.99 (0.27)	5.91 (0.97)	5.83 (0.50)	4.26 (0.27)
	Realigned with MI	5050	106.04 (0.27)	5.89 (0.98)	5.84 (0.49)	4.23 (0.27)
7	Original	484	114.34 (0.37)	6.16 (1.13)	4.39 (0.67)	3.18 (0.33)
	Realigned	447	121.52 (0.31)	8.28 (0.66)	3.82 (0.50)	2.82 (0.23)
	Realigned with MI	4980	120.77 (0.31)	8.38 (0.63)	3.87 (0.49)	2.79 (0.23)
8	Original	382	74.53 (0.47)	3.33 (1.14)	3.99 (0.51)	3.39 (0.26)
	Realigned	321	68.86 (0.45)	3.04 (0.96)	3.54 (0.40)	3.35 (0.24)
	Realigned with MI	4890	65.78 (0.47)	2.98 (0.93)	3.42 (0.40)	3.30 (0.25)

^aData are presented as geometric mean (variance of log-transformed hormone levels).

FSH, follicle-stimulating hormone; LH, luteinising hormone; MI, multiple imputation.

Table 5

Day- and phase-specific associations^a between total dietary fibre intake (per 5 g/day increase) and log serum hormone concentrations among women in the BioCycle Study ($n = 259$), Buffalo, NY, USA, 2005–07

Hormone	Association	Phase ^b	Data	b	SE	[95% CI]
Oestradiol (pg/dL)	Overall	All visits	Original	-0.049	0.019	[-0.087, -0.011]
			Realigned	-0.045	0.019	[-0.083, -0.007]
LH (ng/mL)	Phase-specific	Ovulation (O1, O2, O3)	Realigned with MI	-0.044	0.019	[-0.082, -0.007]
			Original	-0.051	0.025	[-0.100, -0.002]
	Day-specific	Day of LH surge (O2)	Realigned	-0.063	0.024	[-0.111, -0.015]
			Realigned with MI	-0.047	0.022	[-0.090, -0.004]
	Day-specific	Day of LH surge (O2)	Original	-0.062	0.040	[-0.140, 0.016]
			Realigned	-0.110	0.033	[-0.176, -0.045]
FSH (mIU/mL)	Phase-specific	Ovulation (O1, O2, O3)	Realigned with MI	-0.094	0.032	[-0.156, -0.032]
			Original	-0.034	0.018	[-0.068, 0.005]
	Day-specific	Day of LH surge (O2)	Realigned	-0.032	0.017	[-0.065, 0.002]
			Realigned with MI	-0.034	0.016	[-0.066, -0.002]
	Day-specific	Day of LH surge (O2)	Original	-0.019	0.025	[-0.069, 0.031]
			Realigned	-0.061	0.023	[-0.106, -0.016]
Progesterone (ng/mL)	Overall	All visits	Realigned with MI	-0.054	0.022	[-0.097, -0.010]
			Original	-0.077	0.037	[-0.149, -0.005]
	Phase-specific	Luteal phase (L1, L2, L3)	Realigned	-0.086	0.036	[-0.158, -0.015]
			Realigned with MI	-0.066	0.036	[-0.137, 0.005]
	Phase-specific	Luteal phase (L1, L2, L3)	Original	-0.117	0.041	[-0.198, -0.037]
			Realigned	-0.083	0.035	[-0.152, -0.015]
			Realigned with MI	-0.090	0.034	[-0.156, -0.024]

^a All models are adjusted for energy intake, race (White, Black and other), age and vitamin E.

^b Data were realigned to correspond to specific menstrual cycle phases: periovulatory (O1, O2, O3), and early luteal (L1), mid luteal (L2) and late luteal (L3).

b, regression coefficient; CI, confidence interval; FSH, follicle-stimulating hormone; LH, luteinising hormone; MI, multiple imputation; SE, standard error.