

Statistical issues related to modeling menstrual length

Amita K. Manatunga, Brian Schmotzer, Robert H. Lyles, Chanley Small, Ying Guo, Michele Marcus

Menstrual function is of interest to epidemiologists as an indicator of endocrine function and a surrogate for underlying reproductive health. Hormonal patterns affect risk for diseases such as osteoporosis and breast cancer. Menstrual cycle characteristics (e.g., cycle and bleeding length) may reflect a woman's underlying hormonal milieu and potentially offer epidemiologists an immediate, non-invasive measure of reproductive health. Furthermore, menstrual cycle characteristics may indicate important subclinical states, beyond complete reproductive dysfunction.

In reproductive epidemiology, prospective studies of menstrual function often require the statistical modeling of multiple menstrual cycle lengths. Researchers interested in the effect of an exposure on menstrual function may follow women for several months while repeated cycle lengths are recorded. Statistical analyses of these data are not straightforward, and there is relatively little literature exploring statistical issues related to such data.

In this paper, we explore the statistical issues related to modeling repeated menstrual lengths in reproductive studies. Section 1 describes the Mount Sinai Study of Women Office Workers (MSSWOW). This study will provide a framework to discuss issues and approaches to handling repeated menstrual cycle data. Section 2 describes statistical issues related to modeling menstrual lengths. Section 3 describes a simulation study aimed at elucidating the effects of nonstandard features in menstrual function data upon the results of standard linear mixed model analyses. Section 5 discusses the application of these findings to epidemiologic studies of women's reproductive health.

Section 1: A Prospective Study of Reproductive Outcomes (MSSWOW)

The Mount Sinai Study of Women Office Workers (1990-1994) was a prospective cohort study designed to assess the reproductive health of women office workers 40 years of age and younger. 470 women participated in a prospective study of fertility and spontaneous abortion. Participants were interviewed at entry into the study to obtain data on demographics including age and ethnicity, BMI (body mass index, kg/m^2), and reproductive history. During the prospective study, women completed daily diaries containing information on menstrual bleeding, intercourse, birth control use, smoking, alcohol and caffeine consumption. Menstrual cycle length and bleed

length were determined from the daily diary. Cycle length was defined as number of days from the first day of bleeding until the day before the next bleeding episode, and bleed length was defined as the number of days of menstrual bleeding. Women were followed for a median of seven menstrual cycles. 179 women became pregnant during follow-up.

The MSSWOW data serves to illustrate several important statistical issues related to modeling repeated menstrual lengths. Appropriate statistical methods must: 1. Account for the within-woman correlation of repeated cycle lengths; 2. Allow for a non-constant variance across women. Variability in a woman's menstrual lengths reflects biological variation and the measurement error may not be constant across women, a violation of common assumptions in popular statistical models 3. Account for potentially uncompleted last cycles due to administrative censoring. 4. Account for the fact that some women get pregnant during the study period and the rest of their cycle lengths are not observed. Therefore, pregnant women have missing cycle lengths during the study period and the missing mechanism may be informative as opposed to random.

A linear mixed model is a popular statistical method for modeling multiple continuous observations due to its sophistication, flexibility and the availability of software for general use (Laird and Ware, 1982; Diggle, Liang, and Zeger, 1993; Littell et al., 1996). Statisticians and epidemiologists would be tempted to use the linear mixed model for analyzing menstrual length data for these reasons. Most importantly, the linear mixed model can account for the within-subject correlation and can handle time-dependent covariates. Furthermore, the linear mixed model inference procedure can accommodate missing values although the correct inference can only be obtained under the assumption that missing values occurred at random. In this manuscript, we investigate the use of the linear mixed model for modeling menstrual length data. Due to its special features, menstrual data may violate certain assumptions of the standard linear mixed model. For example, variance may be heterogeneous across women; the missingness of cycle lengths may be informative of cycle characteristics. Our purpose here is to assess whether the parameter estimation of covariate effects is impacted by using a linear mixed model with violated assumptions. To achieve this, we conducted several simulation studies. The parameters for the simulation studies were chosen to be roughly representative of the Mount Sinai Study of Women Office Workers (MSSWOW).

3. Statistical Issues:

Because each woman's data consists of multiple menstrual lengths, it appears natural to use the linear mixed model for analysis. Let y_{ij} be the j th menstrual length of the i th woman and let \mathbf{x}_{ij} be the corresponding covariates. We write the model as

$$Y_{ij} = \mu + \alpha_i + \beta' \mathbf{x}_{ij} + e_{ij}, \quad (1)$$

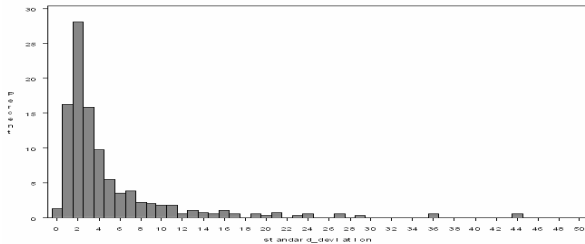
where α_i represents the subject-specific effect and e_{ij} is the random error. We assume that α_i as normally distributed with mean 0 and variance σ_b^2 . The error is assumed to be independent of the subject effect and to follow a normal distribution with mean zero and variance σ_e^2 . Note that the error variance σ_e^2 is typically assumed constant across subjects. The correlation between any two measurements of a randomly selected subject is

$$\frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}, \text{ also known as the intraclass correlation.}$$

As aforementioned, the above linear mixed model has several strengths: (a) it adjusts for within correlation for repeated measurements (b) it can be easily implemented via most popular software such as SAS and Splus, (c) it can accommodate different error covariance structures if needed, and (d) it also can be used to predict subject-specific effects.

However, several statistical issues arise when analyzing repeated menstrual cycle lengths using the standard linear mixed model. Firstly, the last cycle length in a fixed-period menstrual study is usually censored. Exclusion of the last censored cycle may lead to biased estimates (Wang and Chang, 1999). Secondly, an important assumption of this model is that the subject-specific heterogeneity is constant. Figure (1) shows a histogram of women's standard deviations of cycle lengths.

Distribution of Women's Standard Deviation



As seen by this figure, there is considerable variation in standard deviations across women, violating the standard assumption of the linear mixed model. In addition, the

inference procedures that can be implemented via the usual software are based on complete lengths and cannot accommodate censored lengths. When the censored cycles are present, the likelihood can be written to accommodate censored lengths. However, numerical algorithms are required to implement the analysis in this case. Accommodating heterogeneity of within-women variances is also not straightforward and requires extensive modeling procedures with additional assumptions. Finally, the use of the linear mixed model is appropriate when the drop out is random or when the menstrual lengths are missing at random. However, there would be missing menstrual lengths due to pregnancy.

4. Simulation Study

4.1 Exclusion of the last censored cycle

In this simulation study, we examine the effect of the exclusion of the last censored cycle with use of the linear mixed model. We simulated the data from the following model and use a fixed cutoff of 200 days (which was the average follow-up time in the MSSWOW data) as the end of the study period.

$$Y_{ij} = \mu + \alpha_i + \beta_{\text{age}}(\text{age}) + \beta_{\text{smoke}}(\text{smoke}) + e_{ij}, \quad (2)$$

$i=(1, \dots, 400)$, where $\mu = 30, \sigma_b = 5.8, \sigma_w = 7.4$, $\text{age} \sim N(30, 5)$, and $\text{smoke} \sim \text{Bernoulli}(0.3)$. For each simulated dataset, we excluded the last censored cycle and fit the model using the linear mixed model procedure (REML).

Table 1: Exclusion of last censored cycle

	True	Mean	Mean Var	Var*	Coverage
Overall mean	33.1	32.8	4.0	3.9	94
Age	-0.1	-0.10	0.004	0.004	95
Smoke	-0.38	-0.38	0.50	0.54	94
Subject variance	5.8	5.8	0.074	0.070	94
Error variance	7.4	7.4	0.013	0.013	95

Var* stands for Monte Carlo variance

Table 1 summarizes the mean parameter estimates and their variances. There is essentially no bias in the parameter estimates. The mean estimated variances closely approximate the Monte Carlo variances as well.

Obviously, there would be some efficiency loss due to elimination of the last censored cycle, but this loss is expected to be minimal.

4.2 Impact of non-constant variance among subjects

Since the within-subject variance is not typically homogeneous across subjects in menstrual function studies, it would be desirable to determine the effects on estimation under the linear mixed model due to ignoring the heterogeneity. For this purpose, we simulated data from the linear mixed model where the within-woman variance was derived from a non-degenerate distribution. Prior references (e.g., Harlow et al., 2000; Lyles et al., 1999) specifically model such variance heterogeneity. For the purpose of our simulation, we assume that the within-woman variances are lognormally distributed in the underlying population. Specifically, we assume that

$$e_{ij} | \sigma_i^2 \stackrel{iid}{\sim} N(0, \sigma_i^2), \text{ where } \ln(\sigma_i^2) \sim N(\phi, \tau^2)$$

(Lyles et al., 1999). Results of these simulations are presented in Table 2

Table 2: Impact of non-constant variance (lognormal[†])

	True	Mean	Mean Var	Var*	Coverage
Overall mean	33.11	32.63	4.20	4.36	94
Age	-0.1	-0.1	0.0045	0.0047	95
Smoke	-0.38	-0.35	0.53	0.51	96
Subject variance	5.8	5.7	0.084	0.079	94

Var* stands for Monte Carlo variance

[†]Lognormal model for subject-specific variance (Lyles, et al., 1999)

Note that there is no appreciable bias in parameter estimation or standard error estimation regarding covariate effects. The estimated overall mean appears to reflect a slight bias, but not severe enough for an appreciable impact on its confidence interval coverage. Overall, there is little impact on parameter estimation due to ignoring random heterogeneity in within-subject variances. However, there may be some bias if one is interested in predicting individual values. As a separate exercise, we also simulated data where the within variance of a woman depends on covariates via the

general structure $\sigma_i^2 = g(\mu, \beta, \mathbf{x}_i)$. Results are presented in Table 3.

Table 3: Impact of non-constant variance (covariates[†])

	True	Mean	Mean Var	Var*	Coverage
Overall mean	33.11	32.79	4.00	3.91	95
Age	-0.1	-0.1	0.0043	0.0041	96
Smoke	-0.38	-0.37	0.51	0.52	94
Subject variance	5.8	5.8	0.074	0.070	95

Var* stands for Monte Carlo variance

$$^{\dagger} \sigma_i^2 = g(\mu, \beta, \mathbf{x}_i)$$

Again we find negligible bias in parameter estimates. In summary, our simulations suggest that the use of the standard linear mixed model for determining association between covariates and menstrual length is barely affected by the heterogeneity of the within variance. However, further investigation of this point under additional simulation conditions is worthwhile.

4.3 Bias due to dropout resulting from pregnancy

To examine the effect of informative censoring we use a logistic probability model to generate the probability of pregnancy at each cycle. Specifically, the model is given by

$$\log\left(\frac{p_m}{1-p_m}\right) = \lambda_1 \alpha_i + \lambda_2 \alpha_i^2 + \theta_m, \quad (3)$$

where α_i is the true unobserved mean menstrual period length for woman i as defined under model (2), p_m is the probability of pregnancy during the m th cycle, and θ_m represents the baseline risk for pregnancy during cycle m . Thus, we assume that a woman's chance of getting pregnant is quadratically associated with her true length, which reflects U-shaped response curves observed in some studies (Small et al, 2005 and Jensen et al 1999). For the purpose of the simulations, if the probability is greater than 0.9 we classify those individuals as pregnant and drop them out of the study. In this way, the simulation reflects the suspicion that a woman's cycle

characteristics (e.g., average cycle length) are associated with her fecundity. Results are given in Table 4,

Table 4: Bias due to dropouts (pregnancy)

	True	Mean	Mean Var	Var*	Coverage
Overall mean	33.11	33.47	6.72	6.79	94
Age	-0.1	-0.12	0.0072	0.0071	95
Smoke	-0.38	-0.44	0.85	0.87	95
Subject variance	5.8	6.48	0.13	0.09	42
Error variance	7.4	7.36	0.022	0.021	95

Var* stands for Monte Carlo variance

where parameter estimation for each simulated dataset was carried out based on the standard linear mixed model (2).

In this case, there is considerable bias in parameter estimation corresponding to “Smoke”. The between-woman variance component estimate is also biased and its confidence interval demonstrates a low coverage probability. Simulations confirm that as we increase the strength of the association between the average menstrual length and the probability of dropout, the bias becomes more prominent. In summary, depending on the strength of the association between the true length and the dropout, results from the linear mixed models are expected to be biased.

5. Discussion

Menstrual cycle length is frequently used in epidemiologic studies of the effect of environmental exposures on reproductive health. It provides insight into underlying hormonal status and can be measured by questionnaire. These features make assessing menstrual cycle length easier for study participants and less expensive than collection and assay of biological samples. However, within and among woman variation in menstrual cycle length and censoring due to pregnancy make analyses complex.

In this paper, we have explored the use of the linear mixed model for modeling repeated menstrual lengths. The linear mixed model offers great flexibility in analyzing such data, but researchers should interpret the results with caution due to special features often present in such data. Based on the current study, it appears that

the effects of dropping the last (censored) cycle and the effects of ignoring heterogeneity in within-subject variances appear to be relatively minor if interest centers on estimating covariate effects. However, an important issue is that dropout due to pregnancy, which may be closely related to women’s true average cycle lengths, can become a source of noteworthy bias.

Future work will focus on further modeling strategies to adjust for the latter source of bias, as well as joint modeling strategies to allow consistent estimation not only of parameters in the model describing repeated cycle lengths [e.g., model (2)], but also of parameters characterizing the risk of pregnancy in models analogous to (3).

REFERENCES:

- Wang, M.C. and Chang, S.H. (1999). Nonparametric estimation of a recurrent survival function.
- Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* 38:963-974.
- Diggle, P.J., Liang, K.-Y., and Zeger, S.L. (1994). “Analysis of Longitudinal Data”. Oxford University Press: Oxford.
- Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). “SAS System for Mixed Models”. SAS Institute, Inc.: Cary, NC.
- Lyles, R.H., Munoz, A., Xu, J., Taylor, J.M.G., and Chmiel, J.S. (1999). Adjusting for measurement error to assess health effects of variability in biomarkers.
- Harlow, S.D., Lin, X., Ho, M. (2000). Analysis of menstrual diary data across the reproductive life span: Application of the bipartite model approach and the importance of within-woman variance. *Journal of Clinical Epidemiology* 53: 722-733.
- Small CS, Manatunga AK, Klein M, Dominguez CE, Feigelson HS, Marcus M. (2005). Influence of Menstrual Cycle Characteristics on Fertility and Spontaneous Abortion. *Epidemiology*: In press.
- Jensen TK, Scheike T, Keiding N, Schaumburg I, Grandjean P. (1999). Fecundability in relation to body mass and menstrual cycle patterns. *Epidemiology* 10:422-428.