

A hot-deck multiple imputation procedure for gaps in longitudinal data on recurrent events

Roderick J. Little^{1,*},[†], Matheos Yosef², Kevin C. Cain³, Bin Nan¹
and Siobán D. Harlow²

¹*Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109, U.S.A.*

²*Department of Epidemiology, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109, U.S.A.*

³*Department of Biostatistics, University of Washington, Seattle, WA, U.S.A.*

SUMMARY

We consider the analysis of longitudinal data sets that include times of recurrent events, where interest lies in variables that are functions of the number of events and the time intervals between events for each individual, and where some cases have gaps when the information was not recorded. Discarding cases with gaps results in a loss of the recorded information in those cases. Other strategies such as simply splicing together the intervals before and after the gap potentially lead to bias. A relatively simple imputation approach is developed that bases the number and times of events within the gap on matches to completely recorded histories. Multiple imputation is used to propagate imputation uncertainty. The procedure is developed here for menstrual calendar data, where the recurrent events are menstrual bleeds recorded longitudinally over time. The recording is somewhat onerous, leading to gaps in the calendar data. The procedure is applied to two important data sets for assessing the menopausal transition, the Melbourne Women's Midlife Health Project and the TREMIN data. A simulation study is presented to assess the statistical properties of the proposed procedure. Some possible extensions of the approach are also considered. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: event history data; missing data; imputation; menopausal markers

1. INTRODUCTION

This article concerns the analysis of longitudinal data that includes times of recurrent events, where interest lies in variables that are functions of the number of events and the time intervals between events for each individual. Some cases have gaps when the information was not recorded.

*Correspondence to: Roderick J. Little, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.

[†]E-mail: rlittle@umich.edu

Contract/grant sponsor: National Institute of Aging; contract/grant number: AG021543

Discarding the cases with gaps results in a loss of the recorded information in those cases. Other strategies such as simply splicing together the intervals before and after the gap potentially lead to bias. A relatively simple imputation approach is developed that bases the number and times of events within the gap on matches to completely recorded histories. Multiple imputation (MI) is used to propagate imputation uncertainty.

In the problem that motivated this work, the recurrent events are dates of menstrual periods recorded by women, and the substantive question concerns the use of this information to develop markers of menopausal transition. The menopause, or final menstrual period, defines the end of a woman's reproductive life. The transition to menopause occurs over a 5- to 10-year period and is associated with critical changes in women's biology and health status. Yet, unlike puberty, a system for staging reproductive aging has not been definitively established [1]. A staging system would provide a basis for clinical assessment and intervention guidelines and enable women and clinicians to better predict the timing and duration of the transition experience. For example, a staging system would help determine a woman's need for continued contraception, whether bleeding changes require intervention, and when interventions such as targeted hormone therapy (HT) or bone density screening should be initiated. For research, a well-defined staging system would provide a framework for comparisons across studies of midlife women.

In July 2001, the American Society for Reproductive Medicine, the National Institute on Aging, the National Institute of Child Health and Human Development, and the North American Menopause Society cosponsored the Stages of Reproductive Aging Workshop (STRAW), which proposed a staging system [1]. STRAW proposed that, prior to menopause, reproductive life can be divided into reproductive years (three stages) and transition years (two stages, early and late transition). The early- and late-transition stages can be characterized by increasing levels of follicle-stimulating hormone and changes in menstrual bleeding; however, clear criteria for the onset of these stages are still being defined.

The Reproductive Staging (ReSTAGE) project is a collaboration among investigators from four large cohort studies of midlife women: TREMIN [2], the Melbourne Women's Midlife Health Project (MWMHP) [3], the Seattle Midlife Women's Health Study (SMWHS) [4], and the multi-site, multi-ethnic, Study of Women's Health Across the Nation (SWAN) [5]. Its goal is to evaluate the empirically proposed bleeding criteria of the transition stages that served as the basis for the STRAW recommendations (Table I) [4, 6–9]. ReSTAGE compares the reliability, reproducibility, and representativeness of these criteria and assesses how well they predict time to menopause in the four participating cohort studies [10].

A core set of information for the ReSTAGE project in each of these data sets is a history of times when menstrual bleeding episodes occurred, recorded in menstrual calendars kept by the women. The data sets varied in the ages when these calendars were started and in the interval over which they were maintained; these differences are not addressed in this paper, although they do have implications for the staging criteria in Table I.

The key methodological issue addressed here is the fact that all the data sets have gaps in the menstrual calendars, when women failed to record menstrual information. These gaps greatly complicate the calculation of age at the onset of early or late transition, the definitions of which assume that complete menstrual histories are available. One simple approach is to simply discard cases that have any gaps in the menstrual history. The potential selection bias in such an approach might be reduced by weighting the remaining cases by the inverse of the propensity to be complete, estimated by a logistic regression of the indicator for whether a case has a gap on observed

Table I. Proposed bleeding criteria for the onset of the early- and late-transition stages of the menopausal transition as proposed by the Stages of Reproductive Aging Workshop (STRAW) and as defined by four cohort studies of the menopausal transition.

Menstrual bleeding criteria for the onset of			
Early menopausal transition		Late menopausal transition	
STRAW	Other studies*	STRAW	Other studies
Change in cycle length of ≥ 7 days from 'normal' (undefined)	<i>MWMHP</i> (1) >2 cycles < 21 or >35 days in 10 cycles	≥ 2 skipped cycles and ≥ 60 days amenorrhea	<i>MWMHP, SWAN</i> (1) >90 days amenorrhea
	<i>TREMIN</i> (2) Std Dev >6 days in a year		<i>MWMHP</i> (2) Running range >42 days in a 10 cycle sequence
	(3) Std Dev >8 days in a year		<i>TREMIN</i> (3) >60 days amenorrhea
	(4) Cycle >45 days		
	<i>SMWHS</i> (5) ≥ 7 days difference in length of consecutive cycles		<i>SMWHS</i> (4) onset of skipped cycles (\geq twice the modal length)

*MWMHP, Melbourne Women's Midlife Health Project; SMWHS, Seattle Midlife Women's Health Study; SWAN, Study of Women's Health Across the Nation.

covariates. This approach is unappealing as it potentially discards a large amount of information. A better approach is to censor cases with a gap at the age where the gaps begin, but that assumes the censoring mechanism is non-informative and discards recorded information after the gap. It is tempting to ignore the gaps and compute the measures using the intervals with recorded menstruation times. However, discarding a gap implicitly assumes that onset of the transition has not occurred during that gap, and hence tends to push the dates of transition into the future. The extent of bias introduced by this procedure is unknown.

Our approach to gaps is to impute zero or more menstrual bleeding times within the gaps that are plausible given the recorded information. The procedure is a form of hot deck [11] where a match is found to a case with a complete menstrual record, and the number and distribution of bleeding times within the gap are imputed using the information in the matched case. This approach is relatively non-parametric, avoiding the need for building a complex model for the bleeding times, and imputes plausible times within the gap since they are based on real rather than simulated cases.

An advantage of imputation is that, once the data are filled in, all the various potential bleeding events that mark the onset of menstrual transition can be computed unambiguously, with missing data treated the same way for each marker event. Without some form of imputation, the computation of each bleeding marker event needs to be modified to allow for the possibility of gaps in the bleeding histories. It is not obvious how to do this, and the comparability of the assessments is compromised to the extent that the modifications for each marker make different assumptions about the missing values.

A key limitation of imputation, in this as in other contexts, is that information is being added, and hence statistical measures of uncertainty (such as standard errors and confidence intervals)

tend to be underestimated. This limitation is addressed here by MI [11, 12], where multiple data sets are created with different imputations of menstrual bleeding times in the gaps. Imputation uncertainty can then be addressed by applying MI combining rules to analyses conducted on the filled-in data sets.

Specifically, let $\hat{\theta}^{(d)}$ be the estimate of a parameter of interest θ for the d th imputed data set, with associated complete-data variance $W^{(d)}$, $d = 1, \dots, D$. The MI estimate of θ is the average of the estimates from each data set, that is

$$\hat{\theta}_{\text{MI}} = \frac{1}{D} \sum_{d=1}^D \hat{\theta}^{(d)} \quad (1)$$

The MI estimate of the variance of $\hat{\theta}_{\text{MI}}$ is

$$V_{\text{MI}} = \bar{W} + (1 + D^{-1})B \quad (2)$$

where $\bar{W} = \sum_{d=1}^D W^{(d)}/D$ is the average within-imputation variance and $B = \sum_{d=1}^D (\hat{\theta}^{(d)} - \hat{\theta}_{\text{MI}})^2 / (D - 1)$ is the between-imputation variance. Another quantity of interest is the estimated fraction of missing information

$$F_{\text{MI}} = B(1 + D^{-1})/V_{\text{MI}} \quad (3)$$

which measures the fraction of the variance attributable to the fact that data are missing. For refinement of these formulae and combining rules for hypothesis tests, see [11, 12].

The details of the proposed MI method are described in Section 2. Results are illustrated in Section 3 using data from TREMIN and MWMHP. Specifically, the effect of imputation is assessed for two proposed bleeding criteria for, or markers of, the onset of late transition. Section 4 describes a simulation study to assess the statistical properties of the proposed method, and Section 5 discusses some limitations of the MI procedure and potential refinements. Section 6 summarizes the article and places the proposed method in the context of other missing-data approaches.

2. THE IMPUTATION PROCEDURE

We consider data consisting of ages of recurrent events on individuals, with data of some individuals having gaps where the events were not recorded—in our application, the individuals are women and the events are menstrual bleeds, and for concreteness we define the procedure in terms of this application. Consider a gap of length g_i for a woman i , and let a_i be the age of the woman at the start of the gap and $a_i + g_i$ the age at the end of the gap. Thus, it is assumed that the woman had a menstrual bleed at age a_i and a menstrual bleed at age $a_i + g_i$, but an unknown number $n_i \geq 0$ of menstrual bleeds between these ages. In practice, women may have more than one gap, and for simplicity we treat each gap independently in such settings. If $n_i > 0$, let $a_i + \pi_{i1}g_i, \dots, a_i + \pi_{in_i}g_i$ be the (unobserved) times of the bleeding episodes in the gap, where $0 < \pi_{i1} < \dots < \pi_{in_i} < 1$. We call $\pi_i = (\pi_{i1}, \dots, \pi_{in_i})$ the ‘menstrual gap proportions’. Filling in the gap is equivalent to assigning values to n_i , and $\pi_i = (\pi_{i1}, \dots, \pi_{in_i})$ if $n_i > 0$, with the k th bleeding time within the gap imputed at age $a_i + \pi_{ik}g_i$. MI involves repeating this imputation step $Q > 1$ times; thus, for the q th data set, values n_i^q and π_i^q are imputed, and the number of menstrual bleeding episodes and their times

vary across MI data sets. For simplicity, we drop the superscript q for data set and describe the procedure generically for one of these imputations.

Let S be the set of women with complete information about menstrual cycle times. We define the missing data indicator M to take the value $M_i = 0$ for these cases and $M_i = 1$ for cases with a gap to be imputed. The imputation procedure creates a matching set S_i of complete cases that have a pair of menstrual bleeds at ages (say a_{j1} and a_{j2}) that are close to the ages a_i and $a_i + g_i$ that bound the gap. Specifically, for each complete case j , the pair of menstrual bleed ages a_{j1} and a_{j2} (not necessarily consecutive) are computed that minimize

$$d_{ij} = (a_i - a_{j1})^2 + (a_i + g_i - a_{j2})^2$$

Case j is then included in S_i if

$$d_{ij} \leq \text{tol}(g_i)$$

where $\text{tol}(g_i)$ is a pre-specified tolerance that increases with the length of the gap, with coefficients chosen to yield at least five cases in the matching set of each case with a gap, namely

$$\text{tol}(g_i) = 2 + 5(g_i - g_{\min}) / (g_{\max} - g_{\min})$$

where g_{\min} and g_{\max} are the minimum and maximum of all gap lengths.

The imputation for case i then involves the following steps:

- (a) compute the matching set S_i , as described above;
- (b) select a matching case (say j) randomly from the matching set S_i ; for case j , let n_j be the number of menstrual bleeds between a_{j1} and a_{j2} , and if $n_j > 0$, let $0 < \pi_{j1} < \dots < \pi_{jn_j} < 1$ be the gap proportions of the n_j bleeds between a_{j1} and a_{j2} ;
- (c) impute $n_i = n_j$, and if $n_j > 0$, impute $\pi_{ik} = \pi_{jk}$ for $k = 1, \dots, n_j$.

3. APPLICATIONS

3.1. Data

We illustrate the application of the method using representative markers of the menopausal transition and menstrual data from two of the cohort studies participating in ReSTAGE. TREMIN is a publicly available data set from the menstrual calendar study initiated by Treloar *et al.* [2]. It is one of only two data sets that have prospectively recorded menstrual data for individual women throughout their reproductive life span and was initially begun to determine the normal biological variability in menstrual cycle length. Between 1935 and 1939, a total of 1997 female students at the University of Minnesota were enrolled. A second cohort was recruited between 1961 and 1965 and daughters of participants were also invited to participate. Data for this analysis include the subset of 735 women in the original cohort who were still participating at age 35, the baseline age for our analyses of the menopausal transition, and who provided a minimum of 10 consecutive menstrual cycles (Data Tape TRUST998.FINAL supplied in March 1993). The MWMHP enrolled 438 Australian-born women of Anglo-European heritage, aged 45–55 years, who participated in a population-based survey in 1991 and who had menstruated in the prior 3 months and were not using HT. As part

of this longitudinal study of the menopausal transition, women kept records of their menstrual bleeding. Several findings from the MWMHP guided the STRAW recommendations [7, 8]. This analysis includes records from the 216 women who maintained menstrual records through at least 10 cycles.

In both studies, women recorded the days they experienced menstrual bleeding/spotting on a menstrual calendar card that covered one calendar year. In TREMIN, women were asked at the end of every year whether they would continue participating and to complete a short questionnaire with information on pregnancies and medical treatments including HT and contraception. In the MWMHP, participants were followed for up to 8 years with annual assessments, which included reporting of symptom experience and hormone use.

The World Health Organization [13, 14] defines *bleeding episode* as a period of consecutive bleeding days, a *bleeding-free interval* as a period of consecutive bleeding-free days; and a *bleeding segment* as a bleeding episode and the subsequent bleeding-free interval. The term bleeding segment is analogous to the term menstrual cycle, but acknowledges that, in a menstrual calendar, distinguishing between menstrual and non-menstrual (e.g. mid-cycle) bleeds is not straightforward. In ReSTAGE, a single day of bleeding as well as consecutive days of bleeding were coded as a bleeding episode. Bleed-free intervals had to consist of at least 3 days. One or two bleed-free days between two bleeding days were considered to be part of the bleeding episode. Pregnancies and the first three segments after a birth or abortion are coded as non-menstrual intervals. Menopause is defined by the date of a woman's final menstrual period, attributed retrospectively after 12 months of amenorrhea [15].

Women in TREMIN contributed between 4 and 321 segments (median = 145). Women in MWMHP contributed between 4 and 107 segments (median = 36). Menstrual segments ranged in length from 4 to 365 days for TREMIN and from 4 to 325 days for Melbourne.

Many women had some gaps in their menstrual histories ($n = 301$ in TREMIN and 90 in MWMHP) when women failed to return a menstrual card or returned a partially completed card, including cards where women checked that they had not recorded information during a given month. The median number of missing intervals of data is two per woman in TREMIN and two per woman in MWMHP, so there is considerable information to be gleaned from the histories that contain gaps. We constrained the imputation procedure to permit a maximum gap length of 2 years. If a woman had a longer gap, she was censored at the start of the gap and treated as a withdrawal ($n = 33$ in TREMIN).

In our analyses of age at the onset of the late transition, defined by age at the bleeding marker event, women are treated as censored when they withdrew from the study (121 in TREMIN and 48 in MWMHP) or began sustained hormone use or radiation therapy (331 in TREMIN and 59 in MWMHP), or if they were still menstruating at the last observation (MWMHP = 53). Sustained hormone use refers to hormone replacement therapy, hormonal contraception, or other sustained use of steroid hormones. Women who used hormones for a single segment were retained in the data set, with the treated segment coded as a non-menstrual interval. Uncensored women were observed through natural menopause (215 in TREMIN and 42 in MWMHP), surgical menopause (68 in TREMIN and 11 in MWMHP), or endometrial ablation (three in MWMHP).

We consider two markers of the late transition. No bleeding episode for at least 3 months or 90 days is currently the most commonly used definition of the onset of late transition [9]. We set the age at marker to the date of the start of the first observed bleeding episode that initiated a segment of 90 days or longer. Taffe and Dennerstein [7] proposed the running range (the difference

between the longest and shortest cycles observed in a consecutive series of cycles) as an alternative criterion for the onset of the late transition. In the MWMHP, once the running range reached 42 days, fewer than 20 cycles remained before the final menstrual period. For this bleeding criterion, we set the age at marker to the date of the start of the bleeding episode when the running range over a 10-segment sequence exceeded 42 days. These two markers are illustrative of markers defined by the length of an individual segment and by the characteristics of a set of segments, respectively, identification of which may be differentially affected by gaps in the menstrual record.

In addition to calculating the mean age at a bleeding marker event, we apply a varying-coefficient Cox model [16] to investigate the association of a marker event with time to menopause. Previous studies [6] suggest that the association between a bleeding event marking the onset of the menopausal transition and age at menopause may vary with the age at the marker event. The model [16] assumes that the hazard function of menopause changes multiplicatively at time s from the baseline hazard function by a factor of $\exp(\beta(s))$ if the marker happens at time s , that is,

$$\lambda\{t \mid I(t \geq s)\} = \lambda_0(t) \{\exp\{\beta(s)\} I(t \geq s)\} = \begin{cases} \lambda_0(t) \exp\{\beta(s)\} & \text{if } t \geq s \\ \lambda_0(t) & \text{if } t < s \end{cases}$$

where $I(\cdot)$ is an indicator function and $\lambda_0(\cdot)$ is the baseline hazard function. The unknown function $\beta(s)$ is the parameter of interest. Assuming a smooth function, $\beta(s)$ can be approximated by using natural cubic B-splines as the following:

$$\beta(s) = \sum_{k=1}^{K+2} \theta_k B_k(s)$$

where K is the number of interior knots that are usually placed at $100j/(K+1)$ ($j=1, \dots, K$) percentiles of observed marker event times and $B_k(s)$ are B-spline basis functions. For a given set of knots, the coefficients $\theta_1, \dots, \theta_{K+2}$ can be estimated as interaction effects of the marker indicator and spline basis functions at marker time from a time-dependent covariate Cox model, in which the marker indicator is a time-dependent variable taking value 1 if the marker has occurred and 0 otherwise. The set of knots in B-splines is determined by cross-validation. In our data analysis, we use $K=4$. If the marker is right censored, the marker indicator is 0 during the follow-up time period; the marker time can then be assigned an arbitrary value, since it does not affect the parameter estimation.

3.2. Results

We multiply impute the gaps in women's menstrual records using the procedure in Section 2. Figure 1 shows examples of sequences of menstrual cycle lengths including the observed lengths before and after a gap and the imputed lengths within the gap across five imputations for three selected women aged 45–50 years with gap lengths of approximately 100 and 250 days. The figure indicates that the MI procedure fills in the gaps in a plausible way and introduces variability across MI data sets in the number and distribution of menstruation times within the gaps.

We next determine age at marker for each of the two markers of the late transition, 10-segment running range >42 days and no observed bleeding episode for 90 days, in each of the imputed data sets. Table II presents information on the mean and standard deviation of the estimated age

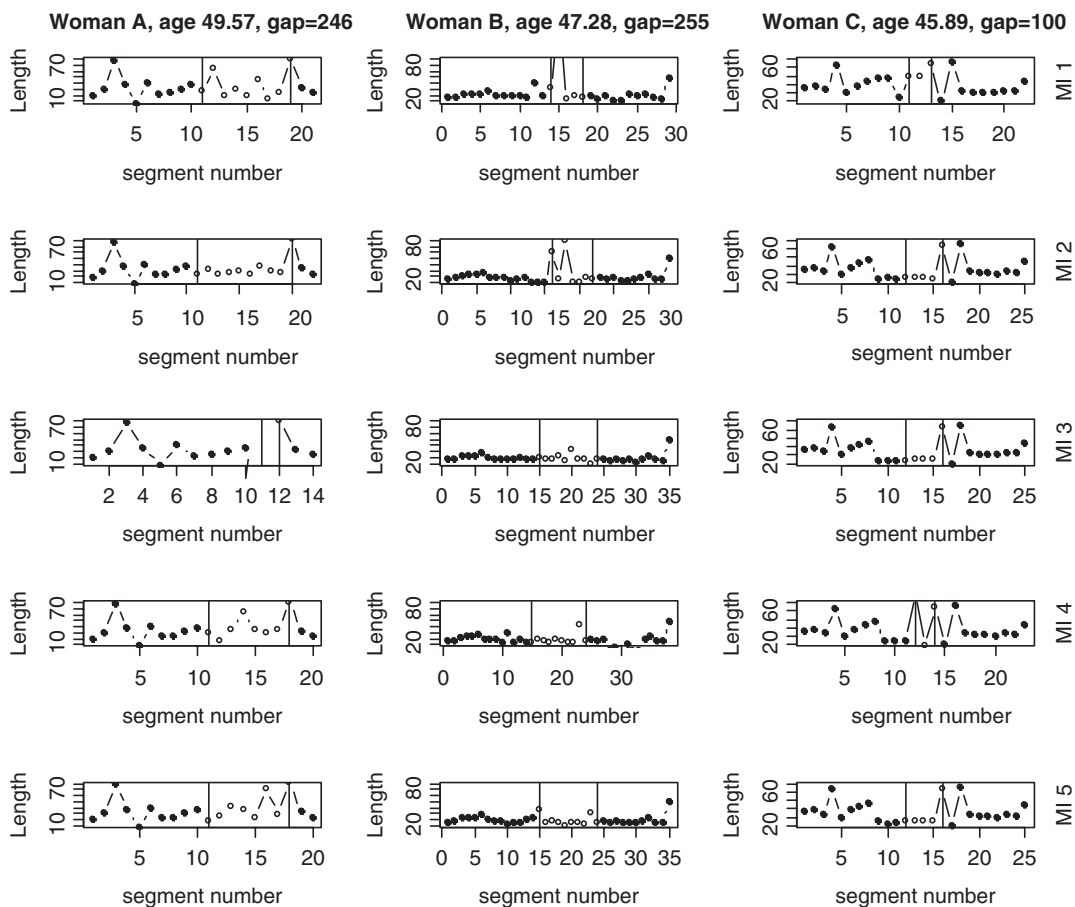


Figure 1. Five imputed sequences of segment length for three selected women with gaps of 246, 255, and 100 days between age 45 and 50 years old. The vertical lines demarcate the gap. Solid circles represent observed events and open circles imputed events. The third plot (MI 3) for woman A has the gap imputed as one segment of length = gap = 246.

The second imputed segment in the gap for woman B has a value of 159.

at marker distributions for the observed data when censored at the first gap, for the observed data when the gap is ignored, for each imputed data set, and for the MI estimate. For each marker, the MI estimate is slightly lower than the observed mean when the gap is ignored but not imputed. The fraction of missing information computed using Equation (3) is very small, indicating that very little information is lost in the menstrual cycle gaps. Note that the MI procedure does not attempt to recover this information, but rather recovers information in the observed cycles of women with gaps that would otherwise be discarded.

Bleeding markers of the onset of the menopausal transition should be associated with the hazard of reaching the final menstrual period. We have previously shown that this hazard changes with age at occurrence of the marker [6] and have developed a varying coefficient approach to

Table II. Mean and standard deviation of age at marker for observed and imputed data.

Marker	Observed, censored at gap	Observed, gap excised	Imputed no. 1	Imputed no. 2	Imputed no. 3	Imputed no. 4	Imputed no. 5	MI summary	Fraction missing
<i>TREMIN</i>									
≥90 days									
N	133	254	255	255	257	254	256		0.00000729
Mean	48.08	48.70	48.68	48.67	48.67	48.66	48.69	48.67	
SD	3.78	3.51	3.50	3.49	3.53	3.55	3.51	3.52	
<i>RR>42 days</i>									
N	189	329	336	332	334	333	336		0.000234
Mean	46.31	46.89	46.78	46.77	46.83	46.68	46.71	46.75	
SD	4.31	4.15	4.2	4.13	4.17	4.29	4.17	4.19	
<i>MWMHP</i>									
≥90 days									
N	36	87	89	89	89	89	89		0.000765
Mean	52.02	51.81	51.79	51.83	51.71	51.76	51.73	51.76	
SD	1.91	1.83	1.80	1.84	1.86	1.80	1.86	1.83	
<i>RR>42 days</i>									
N	57	125	134	132	135	132	131		0.000572
Mean	51.21	51.26	51.20	51.10	51.09	51.12	51.14	51.13	
SD	2.20	1.94	1.92	1.98	1.95	1.90	1.92	1.93	

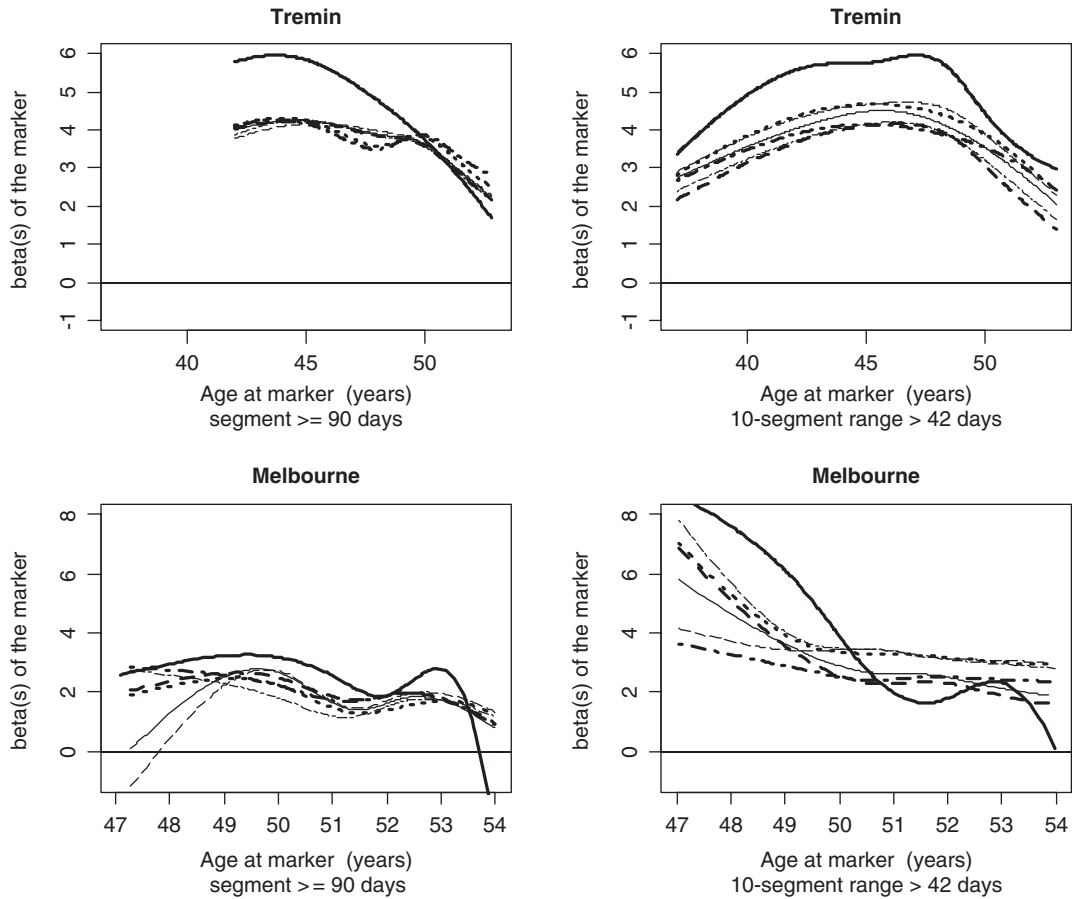


Figure 2. Log-relative hazard of final menstrual period (FMP) by age at marker for unimputed (censored and not censored at first gap) as well as for five imputations for two markers of the late transition in the TREMIN and MWMHP (— unimputed censored at first gap; - - - - unimputed; MI 1; ····· MI 2; - - - - MI 3; ····· MI 4; - - - - MI 5).

model this hazard [16]. Figure 2 describes the log-relative hazard of the final menstrual period by age at marker for the five imputations for the two markers of the late menopausal transition. Figure 3 describes the log-relative hazard of the final menstrual period by age at marker and the 95 per cent confidence intervals using the observed estimates when women are censored at the start of a gap, when the gap is ignored and the MI estimates. The observed hazards when women are censored at the start of a gap are generally higher than the hazards when gaps are ignored or imputed. The observed hazards when gaps are ignored are similar to the imputed hazards and do not differ significantly, although the imputed hazard is slightly higher than the observed hazard across the range of ages for the 10-segment running range >42 days marker.

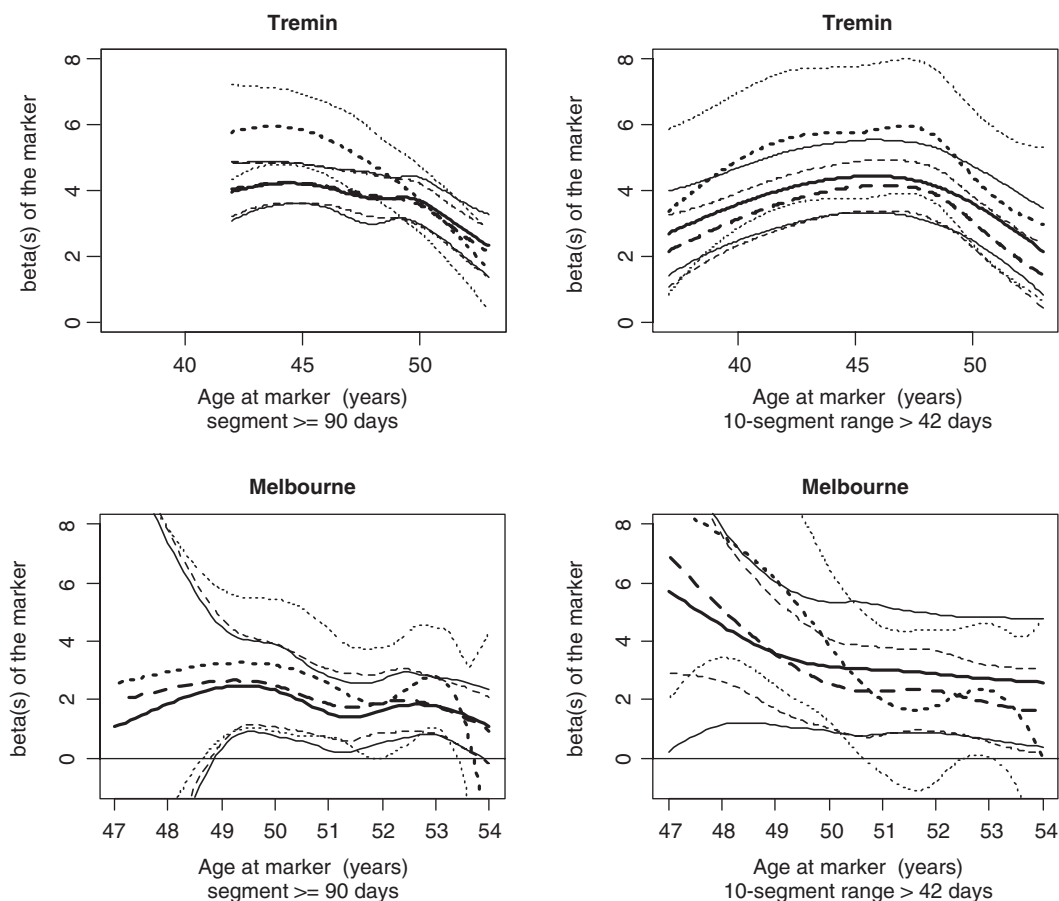


Figure 3. Log-relative hazard of final menstrual period (FMP) by age at marker and 95 per cent confidence bands for unimputed censored and not censored at first gap as well as averaged across five imputations for two markers of the late transition in the TREMIN and MWMHP (— imputed average; --- unimputed; ... unimputed censored at first gap).

4. SIMULATION STUDY

We conducted a simulation study to study further the properties of the MI procedure. We extracted from the TREMIN data the menstrual histories of 190 women with complete information, with the assumption that there are no gaps and no use of hormones. We then created data sets with 30 per cent ($n = 57$) or 60 per cent ($n = 114$) of women having gaps. The number of missing segments in each gap was drawn randomly from a distribution determined empirically by dividing the lengths of the gaps in the incomplete data by the average cycle length of 28 days; the starting segment for each gap was chosen randomly from the set of possibilities. Table III shows summary statistics for the mean age at two of the markers, bleeding interval >90 days (Table IIIa) and 10-segment running range >42 days (Table IIIb). Estimates, empirical bias, and empirical standard deviation are shown

Table III. Simulation study: estimates of mean age at two markers for TREMIN data from data before deletion and imputed data ($N = 190$).

Per cent cases with gaps	Method	Average estimation	Empirical bias	Empirical SD	Added missing data variance
(a) Marker = Bleeding interval > 90 days					
0	Before deletion	48.425	0	0.3686	
30	Single imputation	48.434	0.00871	0.3688	
30	Multiple imputation	48.434	0.00859	0.3699	0.00012
60	Single imputation	48.432	0.00687	0.3669	
60	Multiple imputation	48.430	0.00495	0.3676	0.00048
30*	Single imputation	48.361	-0.0639	0.3956	
30*	Multiple imputation	48.365	-0.0598	0.3856	0.00616
(b) Marker = 10-segment running range > 42 days					
0	Before deletion	46.705	0	0.4046	
30	Single imputation	46.693	-0.0125	0.4093	
30	Multiple imputation	46.693	-0.0127	0.4058	0.00194
60	Single imputation	46.679	-0.0264	0.4134	
60	Multiple imputation	46.672	-0.0334	0.4115	0.00349
30*	Single imputation	46.564	-0.1416	0.4197	
30*	Multiple imputation	46.563	-0.1427	0.4136	0.01031

*Six-month gaps simulated after age 45 in 30 per cent of cases.

Table IV. Simulation study: comparisons of true and imputed number of bleeding intervals in the gaps.
 (a) Six-month gaps simulated after age 38 and (b) six-month gaps simulated after age 48.

Imputed frequency	True					Total
	5	6	7	8	9	
<i>(a) First imputation</i>						
5	0	0	1	0	0	1
6	1	5	8	1	0	15
7	0	5	19	4	0	28
8	0	1	4	5	2	12
9	0	1	0	0	0	1
Total	1	12	32	10	2	57
<i>Second imputation</i>						
5	0	1	0	0	0	1
6	1	4	11	0	0	16
7	0	6	18	8	1	33
8	0	0	3	1	1	5
9	0	1	0	1	0	2
Total	1	12	32	10	2	57

Imputed frequency	True								Total
	1	2	3	4	5	6	7	8	
<i>(b) First imputation</i>									
2	0	0	0	0	0	0	2	0	2
3	0	0	0	0	0	1	0	0	1
4	0	0	1	1	0	0	0	0	2
5	0	0	0	0	0	0	4	1	5
6	0	1	0	2	0	4	4	0	11
7	0	0	1	1	0	3	13	2	20
8	0	0	0	0	2	1	4	2	9
9	1	0	0	0	0	1	1	0	3
10	0	0	0	0	1	1	0	0	2
11	0	0	0	0	1	0	0	0	1
14	0	0	0	0	1	0	0	0	1
Total	1	1	2	4	5	11	28	5	57

Imputed frequency	True								Total
	1	2	3	4	5	6	7	8	
<i>Second imputation</i>									
2	0	0	0	0	0	0	1	0	1
4	0	0	0	0	0	2	2	0	4
5	0	0	0	0	0	0	2	0	2
6	0	0	0	1	1	4	9	1	16
7	1	0	2	2	0	4	7	3	19
8	0	0	0	0	0	1	6	0	7
9	0	1	0	1	3	0	0	0	5
10	0	0	0	0	0	0	1	1	2
12	0	0	0	0	1	0	0	0	1
Total	1	1	2	4	5	11	28	5	57

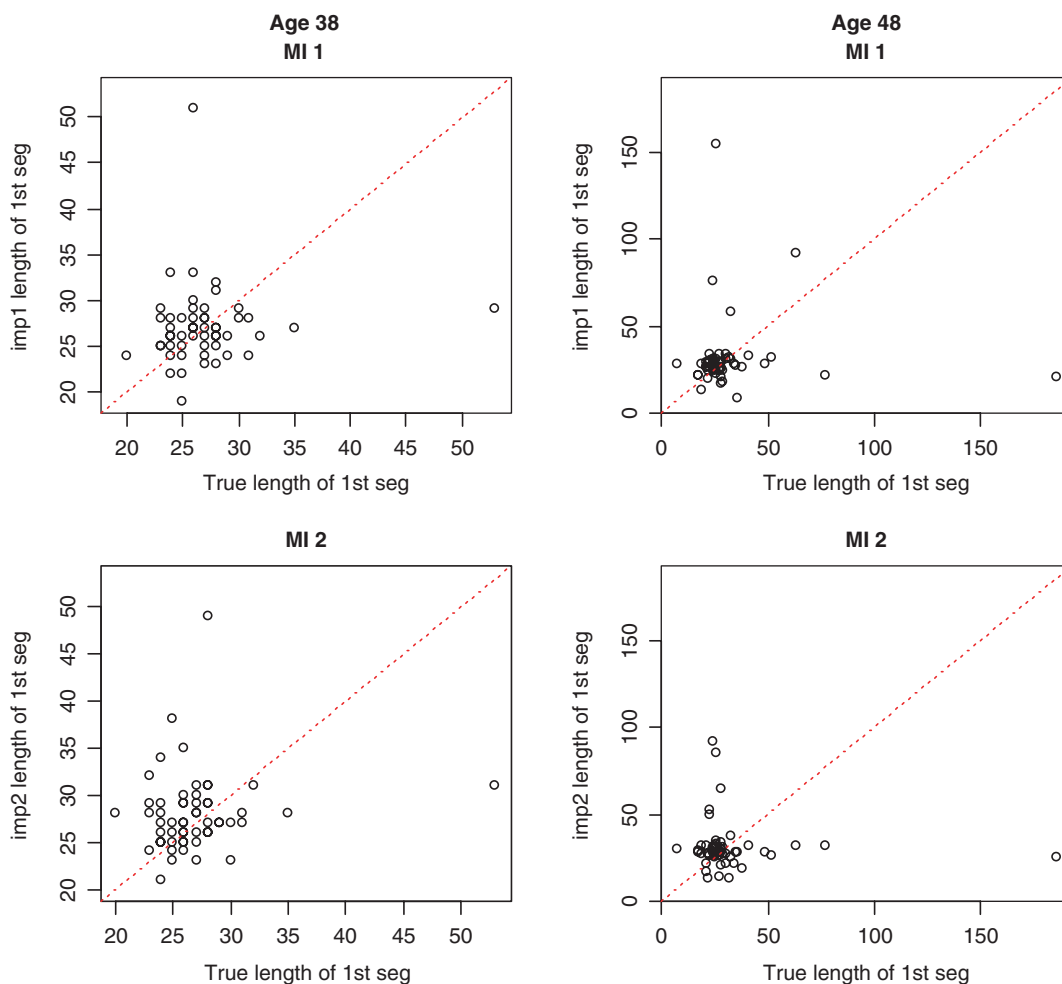


Figure 4. True and imputed first-segment lengths in gaps simulated at ages 38 and 48.

over 100 bootstrap replications of the imputation method, and are shown for: (a) the data before deletion, with empirical standard deviation based on bootstrapping the cases; (b) single imputation using the proposed method; and (c) MI with $M = 5$ using the proposed method. For the latter, we computed the average estimated increase in variance from the missing data. From Table III, we see that the imputation estimates are close to those from the data before deletion, indicating that the procedure has not created much bias. The added variance suggests that the fraction of missing information is very small, so the imputation procedure has recovered significant information in the cases with gaps.

The small fraction of missing information reflects the fact that, for the gap to influence the age at marker, it has to occur around the time when the marker is occurring, which was often not the case for the simulated gaps. We also simulated a situation where this happens more frequently, by creating 6-month gaps in 30 per cent of cases starting at age 45. Results for this case are shown in

the last two rows of Table IIIa and b. The imputation method still yields estimates close to those from the data before deletion. The added variance from missing data is increased, as anticipated given the choice of mechanism, although it is still not very large in this setting.

It is also of interest to compare the imputed information within the gaps with the true information before deletion. Table IV and Figure 4 show two facets of this comparison. Table IV compares the number of actual segments in a gap before deletion with the number of imputed segments, for two of the five sets of MIs. In Table IVa 6-month gaps are simulated starting at age 38, and in Table IVb 6-month gaps are simulated starting at age 48, where menstrual intervals are frequently more variable [2]. The imputation procedure mimics the observed distribution well in Table IVa; in Table IVb also it is quite good, although it tends to impute a few long intervals not seen in the observed data. Finally, Figure 4 compares the true and imputed lengths of the first segment within the gap, indicating that the imputed segments are mimicking reality well.

5. LIMITATIONS AND POTENTIAL REFINEMENTS

For a woman i with a gap, let z_i be other recorded variables, including ages of menstrual episodes before and after the gap, and other covariate information. The imputation method assumes that the data are missing at random (see, e.g. [11]), in the sense that given the information conditioned when creating the matching set, the number and timing of bleeding episodes within the gap, has the same distribution for the case with the gap as for the cases in the matching set. In the matching procedure in Section 2, the only covariate that is conditioned when creating the match is the a_i , the age at the start of the gap. (The existence of a menstrual bleed at or close to $a_i + g_i$ is also conditioned, since it is needed to create comparable information for imputation.) Since changes in the intervals between periods with increasing age are the major feature of interest in developing markers of menopausal transition, a_i is arguably the most important covariate for consideration in matching.

Our procedure might be refined by additional matching on other covariates. For example, one might consider matching on the average length of periods prior to the gap, child-bearing history, or demographic variables such as ethnicity. A limitation of the hot deck is that including additional variables for matching markedly reduces the potential number of matches, and hence the matching tolerance may need to be relaxed if additional matching variables are included. One possibility, mentioned by a referee, is to match on a variable measuring the estimated propensity of a gap, computed by a logistic regression of the indicator for a gap on covariate information [17]. The effects of misspecification of the MI model are more severe when the fraction of missing information is large than when it is small. Thus, the very small fractions of missing information in the analyses in Section 3 suggest that refinements of the method are unlikely to be worth the effort, but that they may be warranted in situations where missing data are more extensive.

A key assumption of the hot-deck procedure is that the probability that $n_i = 0$ for the case with a gap is comparable to the probability that $n_j = 0$ for cases in the matching set. This assumption is violated if the coding of an interval as a gap is related to the incidence of missing periods within the gap—in the extreme, one might assume that only gaps that contain at least one missing period are coded as such, in which case $\Pr(n_i = 0) = 0$. For one of the data sets in the ReSTAGE project, additional information was recorded on whether the gap was thought likely to contain one or more missing periods. It is relatively straightforward to refine our procedure to incorporate this information, as follows.

As before, define the missing data indicator M to take the value $M_i = 0$ for cases with no gap. For the case with the gap, we define the missing data indicator $M_i = m$, where $m = 1$ if there is no evidence in the data to indicate whether $n_i = 0$ or $n_i > 0$, $M = 2$ if there is strong evidence in the data to indicate that $n_i = 0$, and $M = 3$ if there is strong evidence in the data to indicate that $n_i > 0$. The codes $M_i = 2$ and 3 are used for data sets where additional information is collected from the respondent on the nature of the gap.

The imputation procedure first imputes whether $n_i = 0$ or $n_i > 0$. For cases with $m_i = 1$, the odds that $n_i > 0$ is imputed based on the odds estimated from the matching cases in S_i , that is,

$$\frac{\Pr(n_i > 0 | a_i, g_i, z_i, M_i = 1)}{\Pr(n_i = 0 | a_i, g_i, z_i, M_i = 1)} = \frac{\Pr(n_i > 0 | a_i, g_i, z_i, M_i = 0)}{\Pr(n_i = 0 | a_i, g_i, z_i, M_i = 0)}$$

For cases with $M_i = 2$, the complete case odds are multiplied by a prior odds λ_2 set to a value that reflects the prior evidence that the gap does not contain a menstrual bleed; for example, $\lambda_2 = 0.2$

$$\frac{\Pr(n_i > 0 | a_i, g_i, z_i, M_i = 2)}{\Pr(n_i = 0 | a_i, g_i, z_i, M_i = 2)} = \lambda_2 \frac{\Pr(n_i > 0 | a_i, g_i, z_i, M_i = 0)}{\Pr(n_i = 0 | a_i, g_i, z_i, M_i = 0)}$$

For cases with $m_i = 3$, the complete case odds are multiplied by a prior odds λ_3 set to a constant that reflects the prior evidence that the gap does contain a menstrual bleed; for example, $\lambda_3 = 5$

$$\frac{\Pr(n_i > 0 | a_i, g_i, z_i, M_i = 3)}{\Pr(n_i = 0 | a_i, g_i, z_i, M_i = 3)} = \lambda_3 \frac{\Pr(n_i > 0 | a_i, g_i, z_i, M_i = 0)}{\Pr(n_i = 0 | a_i, g_i, z_i, M_i = 0)}$$

The imputation procedure is modified by first imputing whether $n_i = 0$ or $n_i > 0$, by random draws with probability given by the appropriate formula. For cases with $\hat{n}_i = 0$ imputed, the imputation procedure for this gap is complete, since there are no menstrual bleeds in the gap. For cases with $\hat{n}_i > 0$, the menstrual gap proportions π_i are imputed by the hot-deck algorithm described in Section 2, except that only cases with $n_j > 0$ are allowed in the matching set S_i .

An important issue in the development of markers of menopausal transition is the treatment of women who use HT during their reproductive history. In our initial study, these women are excluded since ingestion of hormones disrupts the menstrual cycle and hence measures of menopausal transition. However, the restriction to women not on HT limits the scope of the inference, and in future work we plan to include women on HT in order to study its impact on measures of menopausal transition. One of the approaches to including this information is to treat periods on hormones as gaps in the menstrual history and impute these gaps using the procedure described here. In this application, the potential fraction of missing information may be considerably greater than encountered in the current application, where gaps are simply caused by failure to record the menstrual information.

6. CONCLUSION

We have suggested a relatively simple MI methodology for filling in gaps in histories involving recurrent events, based on hot-deck ideas. The methodology is developed in the context of data on menopausal markers, but it has potential application to any situation where the times of recurrent events are recorded longitudinally, where there are gaps in the records of individual subjects, and

where a method is sought that allows the recorded information in these cases to be included in the analysis. Other examples of recurrent events might include episodes in a chronic illness such as migraine, asthma or depression, or age-sensitive aspects of time-use studies where individuals have intervals where the information is not recorded. In situations where alternative treatments are involved, it would be important to include treatment as a covariate so that information is imputed within each treatment group.

The hot deck is an imputation approach that involves matching each incomplete case to a complete case, called the donor, and imputing the missing variables using the values in the donor case. The method is mainly used to handle item non-response in sample surveys—a classical example is the Current Population Survey hot deck for imputing missing earnings information [18]. A more recent discussion is [19]. To our knowledge, our proposed application of the hot deck to recurrent event data is novel. It also illustrates that the information being transferred from donor to recipient may be a function of the recorded variables. Specifically, here it is the number of events in the gap and the ‘gap proportions’ that are imputed rather than the times of events themselves. This enables the resulting imputed event times to satisfy the constraints implied by the known times that bracket the gaps, creating a seamless imputation. Another somewhat novel feature of the approach is the use of MI by sampling from a match set of potential donors, although the idea of applying the hot deck to create MIs is not new [20, 21].

Useful features of our method are: (a) its simplicity; (b) recovery of the observed information in the records with gaps; (c) availability of simple MI combining rules for analysis of the imputed data, which propagate imputation uncertainty; and (d) the availability of simple measures of imputation uncertainty. As noted above, methods that discard the incomplete records or ignore the gaps are potentially biased and/or lose painstakingly recorded information. From a modeling perspective, an approach that is more principled than that proposed here would be to build a multivariate model for the recurrent event times, conditional on the covariate information, and then estimate the parameters of this model using the incomplete data by maximum likelihood, or multiply-impute the missing values using predictions from this model [11, Chapters 6, 10]. Programs that implement these approaches in repeated-measures settings include PROC MIXED and PROC MI in SAS [22], Iweware [23] and MICE [24]. However, these methods often make distributional assumptions like normality, and are not suitable for recurrent event data where it is the number and timings of events that are missing, rather than missing values of variables measured at particular times. A more appropriate modeling approach here would be to build multivariate event-history models for the repeated events. While this may be a viable option in some settings, it would be challenging to build a suitable model for the data on the menopausal transition studied here, where a key feature is transitions to increasingly irregular spacings between menstrual periods, the timing and nature of which vary across individuals in complex ways. The difficulty in building such non-standard models motivated the simpler approach described here.

ACKNOWLEDGEMENTS

This work was supported by grant number AG021543 from the National Institute of Aging. Data collection for the Melbourne Women’s Midlife Health Project was supported by the Victorian Health Promotion Foundation and the National Health and Medical Research Council of Australia. TREMIN data were kindly provided by the then Tremin Trust, University of Utah, now known as TREMIN and located at Pennsylvania State University. We also thank MaryFran Sowers, an associate editor and a referee, for useful comments on an earlier draft.

REFERENCES

1. Soules MR, Sherman S, Parrott E, Rebar RW, Santoro N, Utian W, Woods N. Executive summary: stages of reproductive aging workshop (STRAW). *Fertility and Sterility* 2001; **76**:874–878.
2. Treloar AE, Boynton RE, Behn BG, Brown BW. Variation of human menstrual cycle through reproductive life. *International Journal of Fertility* 1967; **12**:77–126.
3. Dennerstein L, Dudley EC, Hopper JL, Guthrie JR, Burger HG. A prospective population-based study of menopausal symptoms. *Obstetrics and Gynecology* 2000; **96**(3):351–358.
4. Mitchell ES, Woods NF, Mariella A. Three stages of the menopausal transition from the Seattle Midlife Women's Health Study: toward a more precise definition. *Menopause* 2000; **7**:334–349.
5. Sowers MF, Crawford S, Sternfeld B, Morgenstein D, Gold E, Greendale G, Evans D, Neer R, Matthews K, Sherman S, Lo A, Weiss G, Kelsey J. Design, survey sampling and recruitment methods of SWAN: a multi-center, multi-ethnic, community-based cohort study of women and the menopausal transition. In *Menopause: Biology and Pathobiology*, Wren J, Lobo RA, Kelsey J, Marcus R (eds), vol. 32. Academic Press: New York, 2000.
6. Lisabeth L, Harlow SD, Gillespie B, Lin X, Sowers MF. Staging reproductive aging: a comparison of proposed bleeding criteria for the menopausal transition. *Menopause* 2004; **11**:186–197.
7. Taffe J, Dennerstein L. Menstrual patterns leading to the final menstrual period. *Menopause* 2002; **9**:32–40.
8. Taffe J, Dennerstein L. Time to the final menstrual period. *Fertility and Sterility* 2002; **78**:397–403.
9. Brambilla DJ, McKinlay JB, Johannes CB. Defining the perimenopause for application in epidemiologic investigations. *American Journal of Epidemiology* 1994; **140**:1091–1095.
10. Harlow SD for the ReSTAGE Collaboration (in alphabetical order) Cain K, Crawford S, Dennerstein L, Little R, Mitchell ES, Nan B, Randolph J, Taffe J, Yosef M. Evaluation of four proposed bleeding criteria for the onset of late menopausal transition. *Journal of Clinical Endocrinology and Metabolism* 2006; **91**:3432–3438.
11. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). Wiley: New York, 2002.
12. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.
13. Belsey EM, Farley T. The analysis of menstrual bleeding patterns: a review. *Stochastic Models and Data Analysis* 1987; **3**:125–150.
14. Rodriguez G, Faundes-Latham A, Atkinson LE. An approach to the analysis of menstrual patterns in the critical evaluation of contraceptives. *Studies in Family Planning* 1976; **7**:42–51.
15. World Health Organization. *Research on Menopause in the 1990's*. WHO Technical Report Series. World Health Organization: Geneva, 1996.
16. Nan B, Lin X, Lisabeth L, Harlow SD. A varying coefficient Cox model for the effect of age at a marker event on age at menopause. *Biometrics* 2005; **61**:576–583.
17. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**:41–55.
18. Hanson RH. The current population survey: design and methodology, *Technical Paper No. 40*, U.S. Bureau of the Census, 1978.
19. Marker DA, Judkins, DR, Winglee M. Large-scale imputation of complex surveys. In *Survey Nonresponse*, Groves R, Dillman D, Eltinge J, Little R (eds). Wiley: New York, 2002; 329–342.
20. Little RJA. Missing data in large surveys. *Journal of Business and Economic Statistics* 1988; **6**:287–301 (with discussion).
21. Heitjan DF, Little RJA. Multiple imputation for the Fatal Accident Reporting System. *Applied Statistics* 1991; **40**:13–29.
22. SAS. SAS/STAT software: changes and enhancements, Release 6.07. *Technical Report P-229*, SAS Institute, Inc., Cary, NC, 1992.
23. Raghunathan T, Lepkowski J, VanHoewyk M, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* 2001; **27**:85–95. For associated IVEWARE software see <http://www.isr.umich.edu/src/smp/ive/>
24. Van Buuren S, Oudshoorn CG. Flexible multivariate imputation by MICE. *TNO/VGZ/PG 99.054*, TNO Preventie en Gezondheid: Leiden, 1999. For associated software see <http://www.multiple-imputation.com>