Published in final edited form as:

Menopause. 2013 July ; 20(7): 727–735. doi:10.1097/GME.0b013e3182825ff2.

Classifying Menopausal Stage by Menstrual Calendars and Annual Interviews: Need for Improved Questionnaires

Pangaja Paramsothy, Ph.D. M.P.H.^a, Siobán D Harlow, Ph.D.^a, Michael R. Elliott, Ph.D.^b, Lynda D. Lisabeth, Ph.D. M.P.H.^a, Sybil L. Crawford, Ph.D.^c, and John F. Randolph Jr, M.D.^{a,d}

^aDepartment of Epidemiology, University of Michigan School of Public Health, Ann Arbor MI

^bDepartment of Biostatistics, University of Michigan School of Public Health, Ann Arbor MI

^cDivision of Preventive & Behavioral Medicine, University of Massachusetts School of Medicine
University of Massachusetts School of Medicine, Worcester MA

^dDivision of Reproductive Endocrinology and Infertility, University of Michigan Health System, Ann Arbor MI

Abstract

Objective—To assess agreement between menopausal transition stages defined by annual interview or annual follicle-stimulating hormone measures and menopausal transition stages defined by the monthly menstrual calendar, as well as factors associated with discordance.

Methods—These analyses used daily self-recorded menstrual calendar data from 1996–2006, annual interviews, and annual follicle-stimulating hormone measures. Participants were from 4 study sites of the Study of Women’s Health Across the Nation Boston, southeastern Michigan, Oakland, and Los Angeles, and four racial/ethnic groups: African-American, Caucasian, Chinese, and Japanese. Women who had a defined final menstrual period (FMP) and who never went on hormones were included (n=379). Cohen’s Kappa for 2 by 2 tables were calculated for two definitions of agreement. Logistic regression was used to identify factors associated with discordance.

Results—Poor agreement between annual interview and menstrual calendar data was found for early menopausal transition (Kappa= −0.13, 95%CI: −0.25, −0.02) and late menopausal transition (Kappa= −0.18, 95%CI: −0.26, −0.11). For late stage, Chinese women (OR=2.16, 95%CI= 1.08, 4.30), African-American women (OR=2.39, 95%CI= 1.00, 5.71), and women with a high school education or less (OR=2.16, 95%CI= 1.08, 4.30) were more likely to be discordant. Poor agreement between annual follicle-stimulating hormone measures and menstrual calendars was also found for early menopausal transition (Kappa= −0.44, 95%CI: −0.57, −0.30) and late menopausal transition (Kappa= −0.32, 95%CI: −0.42, −0.23)

Corresponding author: Pangaja Paramsothy, Department of Epidemiology, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, Phone: 734-615-4630, FAX: 734-764-6250, pparams@umich.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

No Conflict of Interest/Financial Disclosure.

DISCLAIMER: *The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH), National Institute on Aging (NIA), the National Institute of Nursing Research (NINR) and the NIH Office of Research on Women’s Health (ORWH).*

Conclusions—New questions need to be developed to accurately identify the start of the menopausal transition and should be evaluated in a multi-ethnic population with varying educational backgrounds.

Keywords

female; menopause; perimenopause; menstrual cycle; questionnaires; bias

INTRODUCTION

In the past decade, the stages of women's reproductive life have become more clearly defined. The 2001 Stages of Reproductive Aging Workshop (STRAW) defined the menopausal transition (MT) as having two distinct perimenopausal stages, early and late. The start of the early MT was defined by increased variability in menstrual cycle length, defined as a change in consecutive cycle lengths of at least 7 days, while the start of the late MT was defined by amenorrhea of at least 60 days¹. These definitions have since been validated by the ReSTAGE collaboration and²⁻⁴ are in the current STRAW +10 recommendations⁵⁻⁸. STRAW also noted that serum follicle-stimulating hormone (FSH) increases during the MT and studies have demonstrated that the mean levels of FSH differ between the different stages of the MT⁹ and the rates of change in FSH level also differ.^{10, 11}

The data that were used to validate the menstrual bleeding markers for early and late MT stages came from prospectively collected menstrual calendars, the preferred method of data collection¹². However, data collection using menstrual calendars is more labor intensive and therefore more costly than studies that use questionnaires. Several cohort studies of midlife women have used annual interviews to classify women's menopausal status, yet few studies have assessed the agreement between information obtained from an annual interview and menstrual calendars in midlife women. The Melbourne Women's Midlife Health Project (MWMHP) found that the interview questions used in their study had low sensitivity for picking up menstrual cycle variability and menses flow variability.¹³ The Seattle Midlife Women's Health Study (SMWHS) compared interview questions inquiring about menstrual cycle irregularity to menstrual calendars and reported weak agreement.¹⁴ Neither of these two studies examined factors that influence agreement. The purpose of this paper was to assess the agreement between MT stages as defined by the annual interview or annual FSH level and MT stages defined by the monthly menstrual calendar in the Study of Women's Health Across the Nation (SWAN) and to examine demographic and lifestyle factors that may influence this agreement.

METHODS

SWAN is a multi-ethnic, multi-site cohort study of middle-aged women. The design of the study has been previously described.¹⁵ Briefly, a cross-sectional screening survey was administered to 16,065 women at seven sites between 1995 and 1997 to assess eligibility for enrollment into the cohort study. Each site recruited Caucasian women and women from one specified minority group (African Americans in Pittsburgh, Boston, southeastern Michigan, and Chicago; Japanese in Los Angeles; Chinese in Oakland; and Hispanic women in Newark). Eligibility for the cohort study included age 42–52 years, an intact uterus, at least one menstrual period and no use of reproductive hormones in the previous 3 months, self-designation as a member of the targeted racial/ethnic group and residence in the geographic area of one of the seven clinic sites, the ability to speak English, Cantonese, Japanese, or Spanish, and the ability to give verbal consent. A total of 3302 women were enrolled into the cohort study at the seven sites. This analysis includes data from the 1950 women who

were enrolled at four study sites (Boston, southern Michigan, Oakland, and Los Angeles) for which cleaned menstrual calendar data were available. Institutional Review Boards at each study site approved the protocol.

Annual study visits began in 1996 and follow-up visits have been conducted since that time. This paper includes data through follow-up visit 10. Each visit consisted of interviewer-administered and self-administered questionnaires that included questions on menstrual characteristics, socio-demographic experience, lifestyle, and medical history. The participants also underwent physical assessments that included a blood draw.

A self-administered monthly menstrual calendar component began in 1996 and continued through 2006. Participants filled out the menstrual calendars daily to capture days where any spotting or bleeding occurred. On the last day of the month women indicated whether no bleeding occurred that month and answered questions about hormone use and gynecological procedures which could affect bleeding. Women were asked to continue filling out and returning the monthly calendar for 2 years after their last menstrual bleed.

Women's menstrual experience was assessed by examining their sequence of menstrual cycle lengths. A menstrual cycle consists of a bleeding episode and a subsequent bleed free interval of at least 3 days. Menstrual cycle length was calculated using bleeding definitions originally developed by the World Health Organization¹² and previously utilized in ReSTAGE analyses^{2, 4}.

A serum sample was obtained at each annual interview, in the morning following an overnight fast, on days 2–5 (days 2–7 from January 1996 through May 1996) of a spontaneous menstrual cycle. Two attempts were made to obtain a day 2–5 sample. If a timed sample could not be obtained, a random fasting sample was taken within 90 days of the anniversary of the baseline visit. FSH assays were conducted using an ACS-180 automated analyzer (Bayer Diagnostics Corp., Norwood, MA). FSH concentrations were measured with a two-site chemiluminometric immunoassay. The interassay coefficient of variation was 12.0%, the intraassay coefficient of variation was 6.0%, and the lower limit of detection was 1.1 IU/liter.

Classification of Reproductive Stage

Menstrual calendar—Bleeding criteria for the onset of early and late MT stages were defined from the calendar data using definitions developed by STRAW/ReSTAGE^{1–4}. We defined the bleeding marker of the early MT as a persistent difference of ≥ 7 days in the menstrual cycle length of consecutive cycles. Persistence was defined as recurrence within 10 cycles of the first variable length cycle. The start date of early MT was then defined as the first day of the first variable length cycle. We defined two bleeding markers for late MT. The first definition, consistent with the SWAN annual interview algorithm, was the first day of the first occurrence of a cycle length of 90 days or greater. The second definition, consistent with the STRAW/ReSTAGE definition, used the first occurrence of a cycle length of 60 days or greater. The final menstrual period (FMP) was defined as the first day of a bleeding segment which was followed by at least 12 months of amenorrhea. For women who had missing calendars during the 12 months of amenorrhea, we accepted the FMP in the menstrual calendar if the date was less than 31 days different from the FMP date identified in the annual interview or if there were only 2 missing calendars during the 12 months of amenorrhea.

Annual Interview

The annual interview ascertained information on menstrual bleeding since the last study visit which was used to define a women's menopause status. The questions, based on the Massachusetts Women's Health Study¹⁶ included the following four questions: "Did you have any menstrual bleeding since your last study visit? Did you have any menstrual bleeding in the last 3 months? What was the date that you started your most recent menstrual bleeding? Which of the following best describes your menstrual periods since your last study visit: have they become farther apart, become closer together, occurred at more variable intervals, stayed the same, become more regular, or don't know?" Early MT stage was defined as the first visit where a woman who had had a menstrual period within the last 3 months reported that the variability of her menstrual periods had increased (either became farther apart, closer together, or with more variable intervals). Late MT stage was defined as the first visit where a woman reported that she had a menstrual period in the last year but had not had a menstrual period in the last three months at the time of the interview. Final Menstrual Period (FMP) was defined retrospectively as the self-reported date of the last menstrual period after at least 12 months of amenorrhea, not due to pregnancy or lactation.

FSH

Based on previous analyses of FSH trajectories using the SWAN population¹⁰, we defined the early MT by FSH level as the first annual visit with a serum FSH concentration between 15 and 29.9 IU/liter. We defined late MT by FSH level as the first annual visit with a serum FSH concentration of 30.0 IU/liter or greater.

Baseline Covariates

Ethnicity was self-defined and categorized as African American, Chinese, Japanese, or Caucasian. Highest education (high school graduate /GED or less than high school versus at least some college) and marital status (single, married, or separated, widowed, or divorced) were assessed at baseline. Economic strain was assessed during the initial screening survey with the question "how hard is it to pay for basics?" and was categorized as very hard, somewhat hard, or not hard. Prior use of female hormone therapy (not oral contraceptives) was assessed during the initial screening survey as well.

Height was measured without shoes using either a metric folding wooden ruler or measuring tape (home and some clinic visits), or a fixed stadiometer (clinic visits). Weight was measured without shoes, and in light indoor clothing, using a portable digital scale (home and some clinic visits) or either a digital or balance beam scale in the clinic. Body mass index (BMI), calculated as weight in kilograms divided by height in meters squared, was categorized as underweight ($<18.5 \text{ kg/m}^2$), normal weight ($18.5\text{--}24.9 \text{ kg/m}^2$), overweight ($25.0\text{--}29.9 \text{ kg/m}^2$), or obese ($\geq 30.0 \text{ kg/m}^2$).

Statistical analyses

Data were analyzed using SAS v. 9.2 (SAS Institute Inc., Carey, NC.). To compare baseline characteristics of eligible women and other participants, Pearson's chi-square or Fisher's exact tests were used to compare proportions, and Student's *t* tests were used to compare means. Analyses were restricted to women who had at least 10 consecutive menstrual cycles recorded in the menstrual calendar and who never start hormone therapy during the study. Since failure to observe transition stages was common among women not observed through the FMP, and the analyses were consistent (data not shown), we present results only for analyses among women with a documented FMP ($n=379$). Women who were already in early MT at the baseline interview – that is, they reported increased variability of menstrual periods in the last year – were excluded from the assessment of early MT as it was not

possible to assess the time of onset of early MT in these women. Similarly, women who had serum FSH concentrations of 15.0 IU/liter or greater at baseline were excluded from assessment of early MT in the comparison with FSH, and women who had serum FSH concentration of 30.0 IU/liter or greater at baseline were excluded from comparisons of both early and late MT in analyses comparing FSH defined MT with MT stage by menstrual calendar.

To examine agreement between the MT stage by annual interview and MT stage by menstrual calendar, the date of entry into each MT stage was determined in the calendar and compared to the MT stage at the next annual interview. If the annual interview MT stage indicated a change in stage from the prior visit, then the annual interview MT stage and the calendar MT stage were determined to have occurred at the same time. If the annual interview MT stage changed at an earlier visit, then the annual interview MT stage was determined to have occurred before the menstrual calendar MT stage. If the annual interview MT stage changed at a later visit then it was determined to have occurred after the menstrual calendar MT stage. If an annual interview MT stage, menstrual calendar MT stage, or both were not observed this was also characterized. The percent in each of these agreement categories was calculated. An example is shown in Figure 1. In this example, the early MT bleeding marker was observed in the menstrual calendar between annual visit 2 and visit 3. If the annual interview at visit 3 observed a change in MT stage from pre MT stage to early MT stage, then the two methods occurred at the same time (0). If the annual interview observed a change at visit 4, then the early MT stage by menstrual calendar is defined as having occurred 1 visit before the annual interview (−1). If the annual interview observed the change at visit 2, then the early MT stage by menstrual calendar is defined as having occurred 1 visit after the annual interview (+1).

Two definitions of agreement were defined for each comparison. Strict agreement was defined as concordant when either the menstrual calendar MT stage and the annual interview MT stage occurred at the same time or both the calendar MT stage and the annual interview MT stage were not observed. Discordance was then defined when the timing of the annual interview MT stage and timing of the menstrual calendar MT stage did not match, or if either the annual interview MT stage or the menstrual calendar MT stage were not observed. A similar definition of strict agreement was used when comparing the calendar to annual FSH level.

When comparing the menstrual calendar to the annual interview, relaxed agreement was defined as the menstrual calendar MT stage and the annual interview MT stage occurred at the same time, the menstrual calendar MT stage occurred 1 visit before the annual interview (−1), or both the menstrual calendar MT stage and the annual interview MT stage were not observed. Relaxed agreement for the menstrual calendar compared to the annual FSH was defined as the menstrual calendar MT stage and the annual FSH MT stage occurred at the same time, the menstrual calendar MT stage occurred 1 visit before the annual FSH (−1), the menstrual calendar occurred 1 visit after the annual FSH (+1), or both the menstrual calendar MT stage and the annual FSH MT stage were not observed.

Cohen's Kappa statistics were calculated for 2 by 2 tables for both strict agreement (Figure 2) and relaxed agreement. The values of kappa can range from −1 to 1, with negative values indicating agreement less than chance, zero indicating chance agreement, and positive values indicating agreement greater than chance.¹⁷ To interpret our Kappa values, we used cutoffs that are similar to the often cited Landis and Koch paper¹⁸: <0.00 poor agreement, 0.01–0.20 slight (poor) agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, 0.81–1.00 almost perfect agreement. Logistic

regression was used to examine baseline factors that influenced discordance, both for strict agreement and relaxed agreement.

RESULTS

Among the 1950 SWAN participants from the four study sites, 1852 (95.0%) enrolled in the calendar study. Participation rates were as follows: Boston 94.0%, south-eastern Michigan 90.8%, Oakland 98.0%, and Los Angeles 97.6%. Of these women, 1339 (72.3%) recorded at least 10 consecutive untreated non-missing cycles and were eligible for this analysis. Of the 1339 eligible women, 379 (28.3%) had their FMP recorded in the calendar study, 23 (1.7%) had a hysterectomy, 381 (28.5%) started taking female hormones, 435 (32.5%) women did not complete the calendar study through FMP or hysterectomy, and 121 (9.0%) were still recording menstrual bleeding at the end of the calendar study. Women who had their FMP recorded were older at study enrollment, and were more likely to be from the Oakland and Los Angeles study sites, thus Chinese or Japanese, than other eligible women. Education, marital status, economic strain, body mass index, and history of female hormone use did not differ at baseline between women whose FMP was and was not recorded. (Table 1)

Comparison of MT stage by menstrual calendar to MT stage by annual interview

For 42.8% of women, the early MT stage by menstrual calendar occurred at the same time as the early MT stage by annual interview. (Table 2) However, a large percentage of women (29.2%) had their early MT stage by menstrual calendar before the onset of early MT stage by the annual interview. A smaller percentage (11.2%) had onset of early MT stage reported in the annual interview before the occurrence of the early MT stage by annual interview. An early MT stage by annual interview, early MT stage by menstrual calendar, or both was not observed in 16.8% of women. Poor agreement was found between the early MT stage by menstrual calendar and the early MT stage by annual interview for both the strict definition (Kappa = -0.13, 95% confidence interval= -0.25, -0.02) and the relaxed definition (Kappa = 0.01, 95% confidence interval= -0.12, 0.13).

Women from the Boston study site were more likely to be discordant (OR=2.17, 95% CI= 1.03, 4.53) than women from Los Angeles when using the strict agreement definition. This association did not persist when using the relaxed definition. Race/ethnicity, education, age, marital status, economic strain, body mass index, and history of female hormone use prior to the study were not associated with discordance for either agreement definitions.

In order to compare similar definitions of late MT stage, the late MT stage by menstrual calendar, defined as first cycle length of at least 90 days, was compared to the late MT stage by annual interview. (Table 2) The two classifications were the same in 13.5% of women. The late MT stage by menstrual calendar occurred earlier than the late MT stage by annual interview in 41.2% of women. Only 1.9% of late MT stage by menstrual calendar occurred after the late MT stage by annual interview. A large percentage of women (21.4%) did not have the late MT stage identified by annual interview but had a late MT stage identified in the calendar, while 12.1% of women did not have their late MT stage identified by either method. Poor agreement was found between the late MT stage by menstrual calendar and the late MT stage by annual interview for both the strict definition (Kappa= -0.18, 95% CI= -0.26, -0.11) and the relaxed definition. (Kappa= -0.05, 95% CI= -0.04, 0.14)

We next examined agreement between the STRAW/ReSTAGE preferred definition for the late MT stage by menstrual calendar (cycle length of at least 60 days) and the late MT stage by the annual interview. The two classifications were the same in 6.1% of the women. Similar to analysis of the 90 day definition, most women had their late MT stage by menstrual calendar occur before the late MT stage by annual interview. (56.2%) A very

small number of women (0.6%) had their late MT stage by menstrual calendar occur after being classified as in late MT stage by the annual interview. The percent of women who did not have late MT stage by annual interview but had a late MT stage by menstrual calendar was 26.6%, while 3.7% had late MT stage by annual interview but no late MT stage observed in the calendar. A small number of women (6.9%) did not have the late MT stage identified by either classification method. Similar to the results for the 90 day definition, poor agreement was found between the 60 day late MT stage by menstrual calendar and the late MT stage by the annual interview for strict agreement (Kappa = -0.08, 95% CI = -0.12, -0.03) and relaxed agreement (Kappa = -0.02, 95% CI = -0.07, 0.03).

We examined factors that influence discordance, using the strict agreement definition, in classification between late MT stage by menstrual calendar and late MT stage by annual interview for the 90 day definition and the 60 day definition. Women from Oakland were more likely to be discordant than women from Los Angeles in both the 90 day definition (OR=2.15, 95% CI = 1.18, 3.93) and the 60 day definition (OR=2.33, 95% CI = 1.04, 5.26). In both the 90 day and 60 day analyses, Chinese women were more likely to be discordant as compared to Caucasian women (late MT definition of 90 days: OR=2.09, 95% CI = 1.04, 4.23; late MT definition of 60 days: OR=2.71, 95% CI = 1.00, 7.36). African-American women were more likely to be discordant as compared to Caucasian women for the 90 day definition (OR=2.39, 95% CI = 1.00, 5.71), but not for the 60 day definition (OR=1.70, 95% CI=0.62, 4.69). Women with a high school education or less were more likely to be discordant in both analyses (the late MT stage of 90 days: OR=2.16, 95% CI = 1.08, 4.30; late MT stage of 60 days: OR=2.92, 95% CI=1.02, 8.40). When examining factors which influence discordance using the relaxed agreement definition, all of the above associations disappeared except for one, women from Oakland were more likely to be discordant than women from Los Angeles in the 60 day definition (OR=1.88, 95% CI = 1.02, 3.45). Age, marital status, economic strain, body mass index, and history of female hormone prior to the study were not associated with discordance in the analyses for the 90 day definition or the 60 day definition for either the strict or relaxed agreement.

Comparison of MT stage by menstrual calendar to MT stage by FSH

A comparison of the early MT stage by menstrual calendar with the early MT stage as defined by the FSH level at the annual visit is shown in Table 3. For 20.8% of the women, the classification of early MT stage by FSH level occurred at the same time as the classification of early MT stage by menstrual calendar. A similar percentage of women (16.1%) had their early MT stage by menstrual calendar occur either before their classification of early MT stage by FSH level or after (15.4%). A large percentage of women (32.2%) had an early MT stage observed in the menstrual calendar but no early MT stage by FSH level, while 10.1% of women had early MT stage by FSH level but no early MT stage by menstrual calendar. A small percentage of women (5.4%) women did not have their early MT stage classified by either method. Poor agreement was found between the early MT stage by menstrual calendar and the early MT stage by FSH level for both strict agreement (Kappa = -0.44, 95% CI = -0.57, -0.30) and relaxed agreement (Kappa = -0.20, 95% CI = -0.33, -0.07). None of the factors examined (study site, race/ethnicity, age, education, marital status, economic strain, body mass index, and history of female hormone use prior to the study) were associated with discordance by either the strict agreement or relaxed agreement definition.

When we compared onset of the late MT based on the menstrual calendar (90 day definition) to onset of the late MT stage by FSH level (Table 3), we observed that many (43.1%) women had their late MT stage by menstrual calendar occur after their classification of late MT stage by FSH level. Only 12.7% of women had their late MT stage by menstrual calendar before their classification of late MT stage by FSH level. A small percentage of

women had a late MT stage observed in the menstrual calendar but not by FSH level (3.3%) or did not have either marker observed (2.0%). However, 20.3% of women had a late MT stage by FSH level but no late MT stage by annual interview. Poor agreement was found between the late MT stage (90 day definition) in the menstrual calendar and the late MT stage by FSH level for both the strict agreement (Kappa= -0.32, 95%CI= -0.42, -0.23) and relaxed agreement (Kappa= -0.12, 95%CI= -0.19, -0.05). Using the strict agreement definition, obese women were less likely to be discordant as compared to women with normal BMI (OR=0.39, 95%CI= 0.20, 0.77). This association became non-significant when using the relaxed agreement definition OR=0.57, 95%CI= 0.32, 1.03). Older women were less likely to be discordant when using the relaxed agreement. For every 1 year increase in baseline age, the OR was 0.91 (95%CI=0.83, 0.99). This association did not hold for the strict agreement definition. Study site, race/ethnicity, education, marital status, economic strain, and history of female hormone use prior to the study were not associated with discordance with either agreement definition.

When we compared the late MT stage by menstrual calendar (60 days definition) to the late MT stage by FSH level (Table 3), 21.0% of women had the late MT stage occurred at the same time by both measures. A similar percentage had their late MT stage by menstrual calendar occur either before (32.0%) or after (36.1%) their late MT stage by FSH level. A small percentage of women (4.2%) had a late MT stage by menstrual calendar observed but no late MT stage by FSH level, (5.6%) had a late MT stage by FSH level but no late MT stage by menstrual calendar observed, or (1.3%) did not have their late MT stage identified by either method. Poor agreement was seen in the comparison of the late MT stage (60 days definition) with the late MT stage by FSH level (Kappa= -0.63, 95%CI= -0.70, -0.56). Unlike the comparison for the 90 days definition, body mass index was not associated with discordance with either definition. Similar to the 90 day definition, older age women were less likely to be discordant when using the relaxed definition. For every 1 year increase in baseline age, the OR was 0.88 (95%CI=0.81, 0.97). Study site, race/ethnicity, education, marital status, economic strain, and history of female hormone use prior to the study were not associated with discordance with either agreement definition.

DISCUSSION

This study assessed agreement between MT stage as defined by a menstrual calendar compared to stage defined by annual interview and by annual FSH level in SWAN. Poor agreement was found between the menstrual calendars and the annual interviews using both a strict definition of agreement as well as a relaxed definition of agreement. Menstrual calendars identified the start of early and late MT earlier than the annual interviews. For the early MT, 29.2 % of women had their status change in the menstrual calendar occur before it was reported in the annual interview. The late MT stage by menstrual calendar 90 days definition occurred earlier in 41.2% of the women and the late MT stage by menstrual calendar 60 days definition occurred earlier in 56.2% of women. When using the strict agreement definition, poor agreement was also found between the menstrual calendar and the annual FSH level using both the strict and relaxed agreement definitions.

Increasing variability in menstrual cycle frequency has been a hallmark description of the onset of the MT.^{4, 19} Using this definition to mark the start of the early MT, our study found poor agreement between the menstrual calendar and the annual interview as the annual interview frequently identified early MT 1–3 years later than the menstrual calendar. Other studies of midlife women have found poor agreement between menstrual calendars and interviews when defining menstrual cycle variability. The MWMHP reported that annual interviews had low sensitivity in detecting menstrual cycle irregularity, defined as change in menstrual cycle frequency.¹³ The SMWHS reported poor agreement between annual

interviews and menstrual calendars in detecting menstrual cycle irregularity (Cohen's kappa = .19).¹⁴

The reasons for poor agreement for the early MT may be due to differences in women's versus researchers' interpretation of menstrual cycle variability. The annual interview question used to define the start of the early MT in SWAN asked a woman to decide if her menstrual periods were farther apart, closer together, or more variable since her last visit. No definition of what farther apart, closer together, or more variable was provided. Therefore, women might not all use the same definition nor use the same definition as researchers. They might only note increased cycle variability when the difference in cycle length is greater than seven days or when differences start to occur more frequently. Our data suggests this may be true given that for approximately one-third of the women the menstrual calendars identified the start of the early MT stage earlier than the annual interviews.

In this study we also found poor agreement between menstrual calendars and annual interviews regarding onset of the late MT stage. We used two definitions for the start of the late MT in the menstrual calendars, a cycle length of at least 90 days and a cycle length of at least 60 days. The 90 days definition gained prominence in the 1990s²⁰ and was easily applied in the clinical setting. Recently, the 60 days definition has been found to be a better marker for the late MT stage, since 9.0% to 21.4% of women have passed their FMP before experiencing 90 days of amenorrhea.² In this analysis, women were more likely to have experienced the late MT bleeding marker of at least 60 days (89.4%) than the late MT bleeding marker of at least 90 days (77.8%).

As with the early MT, the menstrual calendars place the start of the late MT earlier than the annual interview. This finding is not surprising given the structure of the annual interview questions used to identify late MT. The questions were based on the MWHs²¹ which were designed to capture women who were currently experiencing 3 to 11 months of amenorrhea and were thus likely to be classified as post-menopausal at their subsequent visit²⁰. The annual interview failed to detect the late MT stage in 33.5% of the women. All of these women were classified as having changed from early MT directly to being post-menopausal. The annual interview questions ascertained only whether a woman currently had not bled for three months, not whether she had experienced an episode of amenorrhea lasting 90 days or longer in the past year. If in the past year, a woman had had a menstrual cycle that was 90 days or greater but also had a menstrual cycle in the three months preceding the annual interview, she would be classified as being in the early MT by the annual interview definition. However, she would be staged as late MT by the menstrual calendar. A better interview approach to identify the start of the late MT would be to ask about the occurrence of any cycle that was 60 days or greater since the last study visit. A similar approach was used in the SMWHS, which asked questions about skipped periods and found good agreement (Cohen's kappa = .71) between their interview and their menstrual calendar status classification, especially once a definition of skipped period was given to study participants.¹⁴

We found the strict agreement for staging the late MT was influenced by study site, ethnicity, and educational attainment. Chinese women were most likely to have their menstrual calendars and annual interview be discordant. One US study found that Asian women were less likely to accurately recall their last menstrual cycle length; however, the association disappeared once other factors were taken into account including education and income²². Our study found women with a high school education or less were more likely to be discordant. Two studies have found that women with a lower educational attainment were less likely to recall their last menstrual cycle length, however the associations were not statistically significant.^{22, 23} Both of these studies also reported that women with lower

incomes were less likely to accurately recall their last menstrual cycle. When using the relaxed definition of agreement, the majority of the associations we observed disappeared. This suggests that for some women it may take longer to perceive a change in menstrual function.

This study also found poor agreement between menstrual calendars and MT classification by annual FSH levels. Although definitive FSH staging criteria have not been established^{5–8}, recent publications give guidance on appropriate values in the SWAN population¹⁰. However, for the early MT one-third of the women did not have an early MT identified by annual FSH levels suggesting that single annual measurements of FSH levels frequently miss the rise of FSH. Given that FSH is characterized by high variability in this stage^{5–8}, more frequent measurements may be necessary. For the late MT stage of 90 days using the strict agreement definition, we found obese women were less likely to be discordant, consistent with the evidence that obesity affects FSH levels.¹⁰ We also found that for the late MT stage (both 90 day and 60 day definition) using the relaxed agreement definition; older age women were less likely to be discordant. This is not surprising since older age is associated with higher FSH levels independent of menopausal transition stage.²⁴

This study has some limitations. Since women were enrolled into the study between the ages of 42–52 years, left censoring likely occurred such that women may have started the MT before entry into the study. It is possible that the women we determined to be pre-menopausal at baseline interview were already in early MT at the time the study began. Selection bias could also have been a factor. Women who were eligible for this analysis were more likely to be younger, Chinese or Japanese, and to be married as compared to other SWAN participants. Eligible participants were also less likely to be overweight or obese and were less likely to experience economic strain, which could reduce power to detect associations.

CONCLUSION

In conclusion, we found poor agreement between menstrual calendars and annual interviews which use the SWAN framework when staging women during the MT. Accurately identifying MT stage in a timely manner has implications for interventions and healthcare. Bone loss is accelerated in the last couple of years prior to the FMP²⁵, as well as adverse changes in lipid profiles²⁶. Treatment and lifestyle intervention should optimally begin before this time period. Since menstrual calendars are considered the gold standard when measuring menstrual cycle characteristics, these results suggest the need to improve questionnaire based approaches to classification of MT stage. Currently available instruments do not adequately capture the start of the early MT. For the late MT stage, questions that are similar to the skipped periods question used in SMWHS²⁷ are suggested. Given our findings that perception of change in menstrual function differs by ethnicity and educational attainment, questions need to be validated in a multi-ethnic population with varying educational backgrounds. Increasing the frequency of interviews and blood draws may also be warranted. These results also suggest that misclassification of stage may be an important concern in studies that assess change in health status by MT stage based on annual interview classification. Re-analysis of major findings using menstrual calendar based classifications is likely warranted.

Acknowledgments

Clinical Centers: *University of Michigan, Ann Arbor – Siobán Harlow, PI 2011 – present, MaryFran Sowers, PI 1994–2011; Massachusetts General Hospital, Boston, MA – Joel Finkelstein, PI 1999 – present; Robert Neer, PI 1994 – 1999; Rush University, Rush University Medical Center, Chicago, IL – Howard Kravitz, PI 2009 – present; Lynda Powell, PI 1994 – 2009; University of California, Davis/Kaiser – Ellen Gold, PI; University of California,*

Los Angeles – Gail Greendale, PI; Albert Einstein College of Medicine, Bronx, NY – Carol Derby, PI 2011 – present, Rachel Wildman, PI 2010 – 2011; Nanette Santoro, PI 2004 – 2010; University of Medicine and Dentistry – New Jersey Medical School, Newark – Gerson Weiss, PI 1994 – 2004; and the University of Pittsburgh, Pittsburgh, PA – Karen Matthews, PI.

NIH Program Office: *National Institute on Aging, Bethesda, MD – Winifred Rossi 2012; Sherry Sherman 1994 – 2012; Marcia Ory 1994 – 2001; National Institute of Nursing Research, Bethesda, MD – Program Officers.*

Central Laboratory: *University of Michigan, Ann Arbor – Daniel McConnell (Central Ligand Assay Satellite Services).*

Coordinating Center: *University of Pittsburgh, Pittsburgh, PA – Kim Sutton-Tyrrell, Co-PI 2001 – present; Maria Mori Brooks Co-PI 2012; New England Research Institutes, Watertown, MA – Sonja McKinlay, PI 1995 – 2001.*

Steering Committee: Susan Johnson, Current Chair

Chris Gallagher, Former Chair

We thank the study staff at each site and all the women who participated in SWAN.

Funding: The Study of Women's Health Across the Nation (SWAN) has grant support from the National Institutes of Health (NIH), DHHS, through the National Institute on Aging (NIA), the National Institute of Nursing Research (NINR) and the NIH Office of Research on Women's Health (ORWH) (Grants NR004061; AG012505, AG012535, AG012531, AG012539, AG012546, AG012553, AG012554, AG012495). The first author was also supported by the ReSTAGE collaboration which has grant support from the NIA (Grant AG 021543) and the University of Michigan Rackham Graduate School.

REFERENCES

1. Soules MR, Sherman S, Parrott E, et al. Executive summary: Stages of Reproductive Aging Workshop (STRAW). *Fertil Steril*. 2001; 76(5):874–878. [PubMed: 11704104]
2. Harlow SD, Cain K, Crawford S, et al. Evaluation of four proposed bleeding criteria for the onset of late menopausal transition. *J Clin Endocrinol Metab*. 2006; 91(9):3432–3438. [PubMed: 16772350]
3. Harlow SD, Crawford S, Dennerstein L, Burger HG, Mitchell ES, Sowers MF. Recommendations from a multi-study evaluation of proposed criteria for staging reproductive aging. *Climacteric*. 2007; 10(2):112–119. [PubMed: 17453859]
4. Harlow SD, Mitchell ES, Crawford S, Nan B, Little R, Taffe J. The ReSTAGE Collaboration: defining optimal bleeding criteria for onset of early menopausal transition. *Fertil Steril*. 2008; 89(1): 129–140. [PubMed: 17681300]
5. Harlow SD, Gass M, Hall JE, et al. Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. *J Clin Endocrinol Metab*. 2012; 97(4):1159–1168. [PubMed: 22344196]
6. Harlow SD, Gass M, Hall JE, et al. Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. *Menopause*. 2012; 19(4):387–395. [PubMed: 22343510]
7. Harlow SD, Gass M, Hall JE, et al. Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. *Fertil Steril*. 2012; 97(4):843–851. [PubMed: 22341880]
8. Harlow SD, Gass M, Hall JE, et al. Executive summary of the Stages of Reproductive Aging Workshop +10: addressing the unfinished agenda of staging reproductive aging. *Climacteric*. 2012; 15(2):105–114. [PubMed: 22338612]
9. Gracia CR, Sammel MD, Freeman EW, et al. Defining menopause status: creation of a new definition to identify the early changes of the menopausal transition. *Menopause*. 2005; 12(2):128–135. [PubMed: 15772558]
10. Randolph JF Jr, Zheng H, Sowers MR, et al. Change in follicle-stimulating hormone and estradiol across the menopausal transition: effect of age at the final menstrual period. *J Clin Endocrinol Metab*. 2011; 96(3):746–754. [PubMed: 21159842]

11. Sowers MR, Zheng H, McConnell D, Nan B, Harlow S, Randolph JF Jr. Follicle stimulating hormone and its rate of change in defining menopause transition stages. *J Clin Endocrinol Metab.* 2008; 93(10):3958–3964. [PubMed: 18647816]
12. Rodriguez G, Faundes-Latham A, Atkinson LE. An approach to the analysis of menstrual patterns in the critical evaluation of contraceptives. *Stud Fam Plann.* 1976; 7(2):42–51. [PubMed: 936255]
13. Taffe J, Dennerstein L. Retrospective self-report compared with menstrual diary data prospectively kept during the menopausal transition. *Climacteric.* 2000; 3(3):183–191. [PubMed: 11910620]
14. Smith-DiJulio K, Mitchell ES, Woods NF. Concordance of retrospective and prospective reporting of menstrual irregularity by women in the menopausal transition. *Climacteric.* 2005; 8(4):390–397. [PubMed: 16390774]
15. Sowers, M.; Crawford, S.; Sternfeld, B., et al. SWAN: A Multicenter, Multiethnic, Community-Based Cohort Study of Women and the Menopausal Transition. In: L, RA.; K, J.; M, R., editors. *Menopause: Biology and Pathobiology.* San Diego: Academic Press; 2000. p. 175-188.
16. Johannes CB, Crawford SL, Longcope C, McKinlay SM. Bleeding patterns and changes in the perimenopause: a longitudinal characterization of menstrual cycles. *Clinical Consultations in Obstetrics and Gynecology.* 1996; 8:9–20.
17. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med.* 2005; 37(5):360–363. [PubMed: 15883903]
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33(1):159–174. [PubMed: 843571]
19. Treloar AE. Menstrual cyclicity and the pre-menopause. *Maturitas.* 1981; 33(3–4):249–264. [PubMed: 7334935]
20. Brambilla DJ, McKinlay SM, Johannes CB. Defining the perimenopause for application in epidemiologic investigations. *Am J Epidemiol.* 1994; 140(12):1091–1095. [PubMed: 7998591]
21. McKinlay SM, Brambilla DJ, Posner JG. The normal menopause transition. *American Journal of Human Biology.* 1992; 4:37–46.
22. Small CM, Manatunga AK, Marcus M. Validity of self-reported menstrual cycle length. *Ann Epidemiol.* 2007; 17(3):163–170. [PubMed: 16882471]
23. Jukic AM, Weinberg CR, Wilcox AJ, McConnaughey DR, Hornsby P, Baird DD. Accuracy of reporting of menstrual cycle length. *Am J Epidemiol.* 2008; 167(1):25–33. [PubMed: 17928401]
24. Randolph JF Jr, Sowers M, Bondarenko IV, Harlow SD, Luborsky JL, Little RJ. Change in estradiol and follicle-stimulating hormone across the early menopausal transition: effects of ethnicity and age. *J Clin Endocrinol Metab.* 2004; 89(4):1555–1561. [PubMed: 15070912]
25. Lo JC, Burnett-Bowie SA, Finkelstein JS. Bone and the perimenopause. *Obstet Gynecol Clin North Am.* 2011; 38(3):503–517. [PubMed: 21961717]
26. Matthews KA, Crawford SL, Chae CU, et al. Are changes in cardiovascular disease risk factors in midlife women due to chronological aging or to the menopausal transition? *J Am Coll Cardiol.* 2009; 54(25):2366–2373. [PubMed: 20082925]
27. 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(5):334–349. [PubMed: 10993033]

Figure 1.

An Example of the Comparison Between the Annual Interviews (Baseline and Visits) to the Menstrual Calendar. The early menopausal transition (MT) bleeding marker occurs in the menstrual calendar between annual visits 2 and 3. If the annual interview notes a change in menopausal status from premenopausal to early perimenopausal at visit 3 then the two methods are in agreement (0). Instead, if the annual interview notes a change in menopausal status at visit 4, then the early MT stage by menstrual calendar occurs 1 visit prior to the annual interview (-1). If the annual interview notes a change in menopausal status at visit 2, then the early MT stage by menstrual calendar occurs 1 visit after the annual interview (+1).

		Menstrual Calendar	
		Yes	No
Annual Interview	Yes	Same time	Interview stage before calendar stage + Interview stage but no calendar stage
	No	Calendar stage before interview stage + Calendar stage but no interview stage	Neither interview stage nor calendar stage observed

Figure 2.
The 2 by 2 Table for Calculation of Cohen's Kappa Statistic for Strict Agreement.

Table 1

Baseline Demographics of 1339 Women Participating in the SWAN Menstrual Calendar Study with at least 10 Consecutive Untreated Non-Missing Cycles Observed by FMP Status.

	FMP observed N=379	NO FMP observed N=960	P-value ^a
Age at Screener in years , Mean (SD)	46.3 (±2.5)	45.2 (±2.5)	<.01
	n (%)	n (%)	
Study Site			
Michigan	60 (15.8)	241 (25.1)	. <.01
Boston	76 (20.1)	214 (22.3)	
Oakland	107 (28.2)	243 (25.3)	
Los Angeles	136 (35.9)	262 (27.3)	
Race/Ethnicity			
African-American	47 (12.4)	216 (22.5)	<.01
Chinese	72 (19.0)	130 (13.5)	
Japanese	94 (24.8)	132 (13.8)	
Caucasian	166 (43.8)	482 (50.2)	
Language of Baseline Interview			
English	305 (80.5)	863 (89.9)	<.01
Cantonese	32 (8.4)	54 (5.6)	
Japanese	42 (11.1)	43 (4.5)	
Education			
Less than High School	15 (4.0)	36 (3.8)	.99
High School Grad	57 (15.0)	145 (15.2)	
Some College/Vocation	119 (31.4)	302 (31.7)	
College Graduate	91 (24.0)	228 (23.9)	
Post College	97 (25.6)	242 (25.4)	
Missing	0	7	
Marital Status			
Single	62 (16.5)	142 (15.0)	.16
Married	267 (71.0)	635 (66.8)	
Separated	8 (2.1)	32 (3.4)	
Widowed	3 (0.8)	14 (1.5)	
Divorced	36 (9.6)	127 (13.4)	
Missing	3	10	
How Hard Is It To Pay For Basics			
Very Hard	16 (4.3)	73 (7.7)	.08
Somewhat Hard	98 (26.3)	243 (25.6)	
Not Hard	259 (69.4)	632 (66.7)	
Missing	6	12	
Body Mass Index, kg/m ²			
Underweight (<18.5)	13 (3.5)	20 (2.1)	.12
Normal (18.5 –24.9)	219 (58.7)	508 (53.7)	

	FMP observed N=379	NO FMP observed N=960	P-value ^a
Overweight (25.0 –29.9)	69 (18.5)	206 (21.8)	
Obese(30.0)	72 (19.3)	212 (22.4)	
Missing	6	14	
Ever taken hormones at screener	35 (9.4)	115 (12.1)	.16

^aMissing not included in P-value calculation

Table 2
Menopausal Transition (MT) Stage by Menstrual Calendar Compared to MT Stage by Annual Interview among Those with Observed FMP in the Menstrual Calendar

	Early MT Stage ^a (n=250)		Late MT Stage (90 days) (n=379)		Late MT Stage of (60 days) (n=379)	
	n	%	n	%	n	%
No Interview MT stage	7	2.8	81	21.4	101	26.6
Calendar 3+ visits before Interview (-3)	18	7.2	18	4.8	52	13.7
Calendar 2 visits before Interview (-2)	25	10.0	28	7.4	78	20.6
Calendar 1 Visit before Interview ^b (-1)	30	12.0	110	29.0	83	21.9
Calendar and Interview at Same Time ^b (0)	107	42.8	51	13.5	23	6.1
Calendar 1 visit after Interview (+1)	17	6.8	6	1.6	1	0.3
Calendar 2 visits after Interview (+2)	9	3.6	0	0.0	0	0.0
Calendar 3+ visits after Interview (+3)	2	0.8	1	0.3	1	0.3
No Calendar Stage	18	7.2	38	10.0	14	3.7
Neither Calendar nor Interview	17	6.8	46	12.1	26	6.9
Strict Agreement (same time or neither) ^b						
% Agreement ^b	124	49.6	97	25.6	49	12.9
Kappa (95% Confidence Interval) ^b	-0.13 (-0.25, -0.02)		-0.18 (-0.26, -0.11)		-0.08 (-0.12, -0.03)	
Relaxed Agreement (same time, visit before, or neither) ^c						
% Agreement ^c	154	61.6	207	54.6	132	34.8
Kappa (95% Confidence Interval) ^c	0.01 (-0.12, 0.13)		0.05 (-0.04, 0.14)		-0.02 (-0.07, 0.03)	

^aFor comparison of early MT stage by menstrual calendar to early MT stage by annual interview, women who determined to be in early MT stage by baseline interview were excluded. If early MT stage by menstrual calendar occurred at same time or after the late MT stage by menstrual calendar then early MT stage by menstrual calendar was set to missing.

^bThe definition of strict agreement is both methods occur at same time (0) or if the stage was not observed by either calendar or annual interview.

^cThe definition of relaxed agreement is both methods occur at the same time (0), if the calendar occurs 1 visit prior to annual interview (-1), or if the stage was not observed by either calendar or annual interview.

Table 3

MT Stage by Menstrual Calendar Compared to MT Stage by Annual FSH Level among Those with Observed FMP in the Menstrual Calendar.

	Early MT Stage ^a n=149		Late MT Stage (90 days) ^b n=300		Late MT Stage (60days) ^b n=300	
	n	%	n	%	n	%
No MT stage by FSH	48	32.2	10	3.3	12	4.0
Calendar 3 visits before FSH (-3)	6	4.0	7	2.3	17	5.7
Calendar 2 visits before FSH (-2)	6	4.0	8	2.7	23	7.7
Calendar 1 Visit before FSH ^c (-1)	12	8.1	23	7.7	52	17.3
Calendar and FSH at Same Time ^c (0)	31	20.8	56	18.7	63	21.0
Calendar 1 visit after FSH (+1)	13	8.7	59	19.7	49	16.3
Calendar 2 visits after FSH (+2)	6	4.0	23	7.7	26	8.7
Calendar 3 visits after FSH (+3)	4	2.7	47	15.7	29	9.7
No Calendar Stage	15	10.1	61	20.3	25	8.3
Neither Calendar nor FSH	8	5.4	6	2.0	4	1.3
Strict Agreement (same time or neither) ^c						
% Agreement ^c	39	26.1	62	20.7	67	22.3
Kappa and 95% Confidence Interval ^c	-0.44 (-0.57, -0.30)		-0.32 (-0.42, -0.23)		-0.60 (-0.68, -0.53)	
Relaxed Agreement (same time, visit before, or neither) ^d						
% Agreement ^d	56	37.6	138	46.0	164	54.6
Kappa and 95% Confidence Interval ^d	-0.20 (-0.33, -0.07)		-0.12 (-0.19, -0.05)		-0.22 (-0.29, -0.14)	

^aFor comparison of early MT stage by menstrual calendar to early MT stage by FSH, women who determined to be in early MT stage by an FSH level at baseline visit (15.0 IU/liter) were excluded. If early MT stage by menstrual calendar occurred at same time or after the late MT stage by menstrual calendar then the early MT stage by menstrual calendar was set to missing.

^bFor comparisons of late MT stage by menstrual calendar to late MT by FSH, women who determined to be in late MT stage at baseline visit (30.0 IU/liter) were excluded.

^cThe definition of strict agreement is both methods occur at same time (0) or if the stage was not observed by either calendar or FSH.

^dThe definition of relaxed agreement is both methods occur at the same time (0), if the calendar occurs 1 visit prior to annual FSH (-1), if the calendar occurs 1 visit after annual FSH (+1), or if the stage was not observed by either calendar or annual interview.