Practice of Epidemiology

A Method for Longitudinal Prospective Evaluation of Markers for a Subsequent Event

Roderick J. Little*, Matheos Yosef, Bin Nan, and Siobán D. Harlow

* Correspondence to Dr. Roderick J. Little, Department of Biostatistics, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109 (e-mail: rlittle@umich.edu).

Initially submitted February 18, 2010; accepted for publication June 30, 2010.

In this paper, the authors describe a simple method for making longitudinal comparisons of alternative markers of a subsequent event. The method is based on the aggregate prediction gain from knowing whether or not a marker has occurred at any particular age. An attractive feature of the method is the exact decomposition of the measure into 2 components: 1) discriminatory ability, which is the difference in the mean time to the subsequent event for individuals for whom the marker has and has not occurred, and 2) prevalence factor, which is related to the proportion of individuals who are positive for the marker at a particular age. Development of the method was motivated by a study that evaluated proposed markers of the menopausal transition, where the markers are measures based on successive menstrual cycles and the subsequent event is the final menstrual period. Here, results from application of the method to 4 alternative proposed markers of the menopausal transition are compared with previous findings.

biological markers; censored data; life change events; menopause; premenopause; survival analysis

Abbreviations: FMP, final menstrual period; STRAW, Stages of Reproductive Aging Workshop.

Editor's note: An invited commentary on this article appears on page 1388, and the authors’ response appears on page 1391.

It is common in medical and epidemiologic event-history research to seek marker events that are predictive of future events. Examples include staging systems for development of aging processes, staging systems for diseases where transition to a higher stage is associated with increased risk of death, or surrogate markers for survival that allow for more rapid and less costly assessment of alternative treatments. The application that motivated our work concerned assessment of alternative markers of the onset of the menopausal transition, based on the length and variability of menstrual cycles, as indicators of this reproductive life stage and as predictors of age at the final menstrual period (FMP). Unlike puberty, a staging system for reproductive aging has not been definitively established; both the number of stages (1–3) and the precise criteria for defining each stage (1–9) are still being debated. The 2001 Stages of Reproductive Aging Workshop (STRAW) recommendations (3) are the most widely known, and they were revised following our empirical comparative evaluation in 4 large cohorts (6). However, evaluation of the proposed criteria has been limited by the lack of a summary measure capable of comparing several key aspects of multiple proposed markers simultaneously across time (6). In this paper, we present a simple approach that provides an interpretable graphical summary of which proposed marker is more effective at any particular age. The summary measure reflects the differences in the frequency and distribution of age at occurrence of the markers, as well as their ability to predict time to FMP at various ages.

The problem has the following generic features: 1) longitudinal data are available on a sample of individuals from a population for whom the times of intermediate events (markers) are recorded; 2) it is assumed that the markers have not occurred prior to entry into the study; and 3) the ability of the markers to predict the time of a final event F is of interest. The marker m is determined to have occurred at some age ai for individual i, and it is then treated as having...
marker effectiveness. In our comparisons below, we include the proportional hazards model (4, 16). The regression coefficient is treated as a time-varying covariate—for example, in a Cox regression model. The outcome is the time to the event and marker occurrence is binary (3, 16). A more appropriate approach is to apply the methods of survival analysis, where the outcome is a continuous measure and the outcome is binary (12, 13) or varying over time (15). The latter papers extended the notion of receiver operating characteristic curves to the longitudinal assessment of continuous markers. Here we consider the situation where the outcome is a continuous measure and the outcome is binary (12, 13) or varying over time (15). The latter papers extended the notion of receiver operating characteristic curves to the longitudinal assessment of continuous markers. Here we consider the situation where the outcome is a continuous measure and the outcome is binary (12, 13) or the outcome is the time to an event and the marker is continuous and fixed over time (14) or varying over time (15). The latter papers extended the notion of receiver operating characteristic curves to the longitudinal assessment of continuous markers. Here we consider the situation where the outcome is the time to an event and the marker is binary and time-varying, taking the value 0 before the marker event occurs and 1 after it occurs. Comparisons of the sensitivity and specificity of markers, kappa statistics, or receiver operating characteristic curve analysis do not apply naturally in this setting, since the outcome is the time to an event and hence is not binary (3, 16). A more appropriate approach is to apply the methods of survival analysis, where the outcome is the time to the event and marker occurrence is treated as a time-varying covariate—for example, in a Cox proportional hazards model (4, 16). The regression coefficient of the marker in such an analysis provides a measure of marker effectiveness. In our comparisons below, we include a varying-coefficient Cox model (17). However, the size of the coefficient does not reflect the distribution of the occurrence of the marker over time, and differences in the values of the regression coefficients of different markers do not reflect the fact that some markers may tend to occur more frequently or later than others. None of these approaches facilitate comparison of multiple markers simultaneously. Our longitudinal approach takes into account the prevalence of each marker over time, as well as its discriminatory ability, and allows the simultaneous comparison of several markers.

### PROPOSED LONGITUDINAL MEASURE OF MARKER EFFECTIVENESS AND ITS COMPONENTS

The notation used in our method is summarized in Table 1. In a population P of interest, let \( P(a) \) be the set of individuals eligible for marker assessment at age \( a \). For each eligible individual, a marker \( m \) is determined to have occurred or not occurred by age \( a \), based on information available at age \( a \); in our application, the history of menstrual bleeding up to age \( a \). We allow for the possibility that the marker \( m \) is not defined at age \( a \) for a proportion \( 1 - \rho_{ma} \) of the eligible population, because of incomplete information at age \( a \). For example, in our application, one of the markers of menopausal transition is based on the times of 10 consecutive menstrual bleeds prior to the age of interest. The marker is only defined for individuals at ages after the date on which 10 menstrual bleeds have been recorded. For marker \( m \), let \( \pi_{ma} \) be the proportion of definable persons who are positive for the marker—that is, for whom the marker occurred at or before age \( a \)—and let \( 1 - \pi_{ma} \) be the proportion of definable individuals who are negative for the marker at age \( a \).

<table>
<thead>
<tr>
<th>Marker ( m )</th>
<th>A binary indicator of occurrence of an intermediate event predictive of the final event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final event ( F )</td>
<td>The final event being predicted by the markers</td>
</tr>
<tr>
<td>Age ( a )</td>
<td>Age at which the effectiveness of a marker is being assessed</td>
</tr>
<tr>
<td>( \rho_{ma} )</td>
<td>Proportion of the population for which marker ( m ) is defined at age ( a )</td>
</tr>
<tr>
<td>( \hat{\rho}_m )</td>
<td>Estimate of ( \rho_{ma} ) (the circumflex denotes an estimate)</td>
</tr>
<tr>
<td>( \pi_{ma} )</td>
<td>Proportion of definable individuals who are positive for marker ( m ) at age ( a )</td>
</tr>
<tr>
<td>( \mu_a )</td>
<td>Average time from age ( a ) to the terminal event</td>
</tr>
<tr>
<td>( \mu_{ma} )</td>
<td>Average time from age ( a ) to the terminal event, for individuals positive for marker ( m ) at age ( a )</td>
</tr>
<tr>
<td>( \mu_{ma}^- )</td>
<td>Average time from age ( a ) to the terminal event, for individuals negative for marker ( m ) at age ( a )</td>
</tr>
<tr>
<td>( F_{ma}^+ ), ( F_{ma}^- )</td>
<td>Discriminatory functions of time to ( F ) for individuals positive and negative for marker ( m ) at age ( a )</td>
</tr>
<tr>
<td>( \delta_{ma} )</td>
<td>Prevalence factor, defined in equation 4; its estimate is defined in equation 7.</td>
</tr>
<tr>
<td>( \gamma_{ma} )</td>
<td>Overall measure of marker effectiveness, defined in equation 1; its estimate is defined in equation 5.</td>
</tr>
<tr>
<td>( n_a )</td>
<td>No. of eligible individuals in the sample at age ( a )</td>
</tr>
</tbody>
</table>

A binary indicator of occurrence of an intermediate event predictive of the final event is chosen, and the age of the final event is defined as \( v \) if it occurs at age \( v < A \); otherwise it is defined as \( A \) (10, 11).

Previous work has focused on situations where the marker is a continuous measure and the outcome is binary (12, 13) or the outcome is the time to an event and the marker is binary and fixed over time (14) or varying over time (15). The latter papers extended the notion of receiver operating characteristic curves to the longitudinal assessment of continuous markers. Here we consider the situation where the outcome is the time to an event and the marker is binary and time-varying, taking the value 0 before the marker event occurs and 1 after it occurs. Comparisons of the sensitivity and specificity of markers, kappa statistics, or receiver operating characteristic curve analysis do not apply naturally in this setting, since the outcome is the time to an event and hence is not binary (3, 16). A more appropriate approach is to apply the methods of survival analysis, where the outcome is the time to the event and marker occurrence is treated as a time-varying covariate—for example, in a Cox proportional hazards model (4, 16). The regression coefficient of the marker in such an analysis provides a measure of marker effectiveness. In our comparisons below, we include the proportional hazards model (4, 16). The regression coefficient is treated as a time-varying covariate—for example, in a Cox regression model. The outcome is the time to the event and marker occurrence is binary (3, 16). A more appropriate approach is to apply the methods of survival analysis, where the outcome is the time to the event and marker occurrence is treated as a time-varying covariate—for example, in a Cox proportional hazards model (4, 16). The regression coefficient of the marker in such an analysis provides a measure of marker effectiveness. In our comparisons below, we include a varying-coefficient Cox model (17). However, the size of the coefficient does not reflect the distribution of the occurrence of the marker over time, and differences in the values of the regression coefficients of different markers do not reflect the fact that some markers may tend to occur more frequently or later than others. None of these approaches facilitate comparison of multiple markers simultaneously. Our longitudinal approach takes into account the prevalence of each marker over time, as well as its discriminatory ability, and allows the simultaneous comparison of several markers.

### PROPOSED LONGITUDINAL MEASURE OF MARKER EFFECTIVENESS AND ITS COMPONENTS

The notation used in our method is summarized in Table 1. In a population \( P \) of interest, let \( P(a) \) be the set of individuals eligible for marker assessment at age \( a \). For each eligible individual, a marker \( m \) is determined to have occurred or not occurred by age \( a \), based on information available at age \( a \); in our application, the history of menstrual bleeding up to age \( a \). We allow for the possibility that the marker \( m \) is not defined at age \( a \) for a proportion \( 1 - \rho_{ma} \) of the eligible population, because of incomplete information at age \( a \). For example, in our application, one of the markers of menopausal transition is based on the times of 10 consecutive menstrual bleeds prior to the age of interest. The marker is only defined for individuals at ages after the date on which 10 menstrual bleeds have been recorded. For marker \( m \), let \( \pi_{ma} \) be the proportion of definable persons who are positive for the marker—that is, for whom the marker occurred at or before age \( a \)—and let \( 1 - \pi_{ma} \) be the proportion of definable individuals who are negative for the marker at age \( a \).

Let \( \mu_a \) be the average time from age \( a \) to \( F \) for individuals eligible at age \( a \). Similarly, let \( \mu_{ma}^+, \mu_{ma}^- \) be the average times from age \( a \) to \( F \) for eligible individuals for whom the marker \( m \) is defined and who are positive (+) and negative (−) for
marker \( m \) at age \( a \), respectively. Our measure of the effectiveness of marker \( m \) at age \( a \) is

\[
\hat{e}_{ma} = 2\hat{p}_{ma} \pi_{ma} (1 - \pi_{ma}) \times (\mu_{ma} - \mu_{ma}^+),
\]

and it is interpreted as the average prediction gain in the population from knowing that the marker is positive or negative at age \( a \). The change in the predicted time to the final event from knowing that marker \( m \) is positive at age \( a \) is \( \mu_{ma} - \mu_{ma}^+ \), the difference in expected time to the final event and the expected time to event for individuals positive for the marker. Similarly, the change in predicted time to the final event from knowing that the marker is negative at age \( a \) is \( \mu_{ma} - \mu_{a} \). When the marker is not defined, the change in predicted time is 0, since no information is gained. (Statistically, there may be information about the expected time to \( F \) when the marker is not defined, but we assume that this information is not available or, if available, is not exploited in the analysis.) The average prediction gain in the eligible population from knowing that the marker is positive or negative at age \( a \) is obtained by summing the prediction gains, weighted by their respective proportions in the population, yielding

Prediction Gain = \( \hat{p}_{ma} \pi_{ma} \times (\mu_{a} - \mu_{ma}^+) \)
+ \( \hat{p}_{ma} (1 - \pi_{ma}) \times (\hat{\mu}_{ma} - \mu_{a}) \)
+ \( (1 - \hat{p}_{ma}) \times 0 \) (positive for marker)
+ \( (\text{unknown for marker}) \)

Substituting \( (\mu_{a} - \mu_{ma}^+) = (1 - \pi_{ma}) \times (\mu_{ma} - \mu_{ma}^+) \) and \( (\hat{\mu}_{ma} - \mu_{a}) = \pi_{ma} \times (\hat{\mu}_{ma} - \mu_{ma}^+) \) in this expression leads (with simple algebra) to the effectiveness measure \( \hat{e}_{ma} \) in equation 1.

The measure \( \hat{e}_{ma} \) is a product of 2 components:

\[
\hat{e}_{ma} = \hat{\delta}_{ma} \times \gamma_{ma},
\]

where we call the first component

\[
\hat{\delta}_{ma} = \mu_{ma} - \mu_{ma}^+
\]

the discriminatory ability of marker \( m \) at age \( a \) and we call the second component

\[
\gamma_{ma} = 2\hat{p}_{ma} \pi_{ma} (1 - \pi_{ma})
\]

the prevalence factor of the marker \( m \) at age \( a \). The discriminatory ability measures the extent to which occurrence of the marker improves prediction of time to the final event. The prevalence factor reflects the proportion of individuals for whom the marker is defined and positive at age \( a \). It reflects the intuition that, for a given level of discriminatory ability, a marker is more useful in the aggregate when it divides the population approximately equally (e.g., 40% or 60% of the population are positive) than when it divides the population unequally (e.g., 2% or 98% of the population are positive). Specifically, the prevalence factor attains a maximum value of one-half when 50% of individuals are positive and 50% are negative, which is also the distribution of marker prevalence with the highest variance; it declines as the proportion positive tends to 0 or 1.

The inclusion of this prevalence factor (equation 4) is a distinctive feature of our proposed measure (equation 1). Standard regression-based measures of marker effectiveness, such as the regression coefficient of marker incidence in a Cox proportional hazards model, measure discriminatory ability but do not reflect the prevalence of the marker at any given age. Inclusion of the prevalence factor redresses the situation where a regression coefficient is large for a variable that has little impact on aggregate prediction because it has very low (or high) prevalence.

Estimates of the quantities in equations 1, 3, and 4 can be obtained from a suitable, ideally random, sample of the population. Let \( n_{ma} \) be the number of eligible individuals in the sample at age \( a \), \( \hat{p}_{ma} \) be the proportion of these individuals for whom the marker is defined, \( \pi_{ma} \) be the proportion of definable individuals who are positive for the marker (i.e., for whom the marker occurred at or before age \( a \)), and \( 1 - \pi_{ma} \) be the proportion of definable individuals who are negative for the marker.

Let \( \hat{\beta}_{ma} \) be the estimated average time from age \( a \) to \( F \) for individuals who are eligible at age \( a \). Similarly, let \( \hat{\beta}_{ma}^-; \hat{\beta}_{ma}^+ \) be the estimated average times from age \( a \) to \( F \) for eligible individuals who are positive (+) and negative (−) for marker \( m \) at age \( a \), respectively. If the time to \( F \) is not censored and hence recorded for all cases in the sample, these estimates could be the respective sample means, but if some times to \( F \) are censored, a method is needed to estimate the times for censored individuals, as is the case in our application. We then estimate the marker effectiveness of marker \( m \) at age \( a \) as

\[
\hat{e}_{ma} = \hat{\delta}_{ma} \times \hat{\gamma}_{ma},
\]

the product of the estimated discriminatory ability

\[
\hat{\delta}_{ma} = \hat{\beta}_{ma}^- - \hat{\beta}_{ma}^+
\]

and the estimated prevalence factor

\[
\hat{\gamma}_{ma} = 2\hat{p}_{ma} \hat{\pi}_{ma} (1 - \hat{\pi}_{ma}).
\]

We compute the standard errors of equations 5–7 by simple bootstrapping of individuals in the sample.

**APPLICATION: ASSESSMENT OF LATE MARKERS OF MENOPAUSAL TRANSITION**

In 2001, the STRAW recommendations stipulated that reproductive life could be characterized as including 2 menopausal transition stages, early and late (5, 6). Entry into early menopausal transition was characterized by increasing levels of follicle-stimulating hormone and increasing variability in menstrual cycle length. Entry into late menopausal transition was characterized by continued elevation of follicle-stimulating hormone levels and the occurrence of skipped cycles or amenorrhea. The STRAW
recommendations have gained acceptance, but debate remains as to how best to define bleeding markers of the onset of each stage (6).

We applied our methods to 3 menstrual bleeding markers proposed for defining the onset of late menopausal transition which were considered in the empirical evaluation of the STRAW recommendations (6). In these definitions, a segment is the time interval (in days) between 2 menstrual bleeds, defined precisely in the article by Harlow et al. (7). These markers reflect the notion that entrance into late menopausal transition is characterized by segments of increased or more variable length:

a) the first segment of at least 90 days (D90) (5),

b) the first segment of at least 60 days (D60) (8), and
c) the first instance of a running range of more than 42 days (RR10) (9). The running range is computed as the difference between the maximum and minimum lengths of 10 consecutive segments.

Note that point c requires data on 10 successive segments, and hence is not defined until data on 10 segments have been recorded. This motivates the definitions of $p_{ma}$ and its estimate $\hat{p}_{ma}$ given above.

We also include, for illustration, 1 proposed marker of early menopausal transition:

d) persistent >7-day difference in consecutive segment lengths (DIFF7p). This marker is defined as the first segment whose length is at least 7 days greater than that of the previous segment, when at least this magnitude of difference between consecutive

cycles is observed again within the subsequent 10 segments (1).

Our interest is in assessing empirically the relation between these events and the date of the FMP, defined retrospectively as the menstrual period that is followed by 12 months of amenorrhea. When assessing markers a–d, it is assumed that the first occurrence of these events did not take place before a woman was enrolled in the study. For methods of adjusting for left-truncation and censoring from late entry into a study, see the article by Cain et al. (18).

We compared markers using data from the TREMIN study (19), which prospectively recorded the menstrual cycles of students enrolled at the University of Minnesota from 1935 to 1939. This analysis includes records from 726 women who were enrolled by age 25 years, participated for at least 5 years, and were still participating and not using hormones at age 35 years, the baseline age for these analyses. (The data tape TRUST998FINAL was supplied by the TREMIN investigators in 1993.)

All untreated bleeding segments of each eligible woman observed from age 35 years through the FMP or censoring due to hysterectomy, withdrawal, or hormone use were included, with women contributing up to 322 segments. A nonparametric approach was applied to estimate the times to FMP for censored cases. Specifically, let $F_{ma}^+$, $F_{ma}^-$ denote the distribution functions of the time to $F$ for women positive and negative for the marker $m$ at age $a$, respectively, and let $\hat{F}_{ma}^+$, $\hat{F}_{ma}^-$ denote the Kaplan-Meier estimates of these distributions, computed using the set of women for whom the marker is defined at age $a$. Then

Approximately half of the women had gaps in their menstrual histories, with a median of 2 gaps per woman. Gaps in the reported menstrual histories were multiply imputed using a hot deck method described elsewhere (22). Since results from imputed data differed little from results from unimputed data, we present only the latter here. Gaps of less than 4 years were ignored, and women with gaps of 4 years or more were censored at the gap.

The median ages at occurrence of DIFF7p, RR10, D60, and D90 were 41.5, 48.1, 48.1, and 50.0 years, respectively. Thus, RR10 and D60 occurred at similar times; D90 occurred about 2 years later, on average; and the marker of early transition, DIFF7p, occurred 6.5 years earlier (Figure 1). Harlow et al. (7) modeled the hazard function for the FMP using a varying-coefficient Cox model (17) that incorporated censored and uncensored women. The log relative hazard of having one’s FMP as a function of age is plotted for each marker in Figure 2. At any given age, the log relative hazards were similar for each of the 3 late transition markers, but the log relative hazard for DIFF7p was significantly lower than that for the other markers, reflecting the expected finding that the early marker is less predictive of the FMP than the late markers.

Figure 3. Measures of A) marker effectiveness, B) prevalence factor, and C) discriminatory ability for 4 markers of the final menstrual period, by age. —, D60 (segment $\geq 60$ days); - - -, D90 (segment $\geq 90$ days); - - - - , RR10 (10-segment running range); : - : - : - , Diff7p. See text for detailed definitions of marker events.

For these estimates to be well-defined, the longest times in the data set in the positive and negative marker groups have to be events, and not censored. If the last observation time happens to be a censoring time, we put the remaining probability mass on that time point. For these estimates to be consistent, the support of the censoring time needs to contain the support of age at FMP (20, 21). This is probably the case for this data set, since the TREMIN study has long follow-up, with the last observation time often being the FMP (19).
In Figure 3, part A, the estimated marker effectiveness (equation 5) is plotted as a function of age for all 4 markers. From these data, we see that 1) the early marker DIFF7p is more effective than the late markers before age 45 years and less effective after age 45 years, confirming empirically the early-versus-late designation; 2) RR10 is similar to D60 before age 49 years and inferior to D60 after age 49 years; and 3) D90 is similar to D60 after age 52 years but inferior to D60 at earlier ages. Overall, D60 seems to be the best of the 3 late markers in this data set (6).

Further insight into marker performance can be gained by plotting the components of marker effectiveness, prevalence factor and discriminatory ability, against age. Parts B and C of Figure 3 show these plots for the 4 markers. From these plots, we note that D90 has a very small prevalence factor before age 47 years, since the proportion of persons positive for this marker is small for these ages. The estimated discriminatory ability is actually higher for D90 than for D60 in this age range, but it is highly variable because of sampling error. Discriminatory ability tends to increase with age for D60 and decline with age for D90. The superiority of D60 as an overall measure between ages 42 and 52 years seems mainly attributable to the greater prevalence factor for D60 in this age range, since the proportion of women positive for D60 is higher for this less stringent criterion. Since the choice of 60 days and 90 days in these markers is somewhat arbitrary, it would be possible to extend this comparison to a wider set of markers (e.g., D45, D50, ... , D90) to seek an optimal choice.

To conduct statistical tests and calculate confidence intervals for the differences in Figure 3A, we compute bootstrap standard errors for the difference in marker effectiveness between pairs of markers at each age. The associated 95% confidence intervals, given as the difference plus or minus 2 standard errors, are shown for the markers RR10 and D60 in Figure 4. Here the standard errors are computed using 500 bootstrap samples of women in the data set. From this figure, we see that confidence intervals for the differences in marker effectiveness include 0 at early ages but are positive for some ages beyond 50 years, suggesting an advantage for D60 for later ages.

The plot of discriminatory ability in Figure 3C shows the differences in predicted mean time to FMP for women who are negative and positive for the marker at each age. The predicted means for the positive and negative marker groups can also be plotted individually against age, as shown in Figure 5 for the D60 marker; this can be viewed as summarizing plots of the distributions of each group by age, as implemented by Lisabeth et al. (8). In interpreting Figure 5, note that 1) the time to FMP inevitably declines as women get older, 2) the composition of the positive and negative marker groups is changing over time, and 3) the plotted values do not account for the changes in the prevalence of the marker over age. Points 2 and 3 motivate our plot of the combined measure of marker effectiveness (Figure 3A).

**DISCUSSION**

Here we have proposed a longitudinal measure of marker effectiveness (equation 1) for longitudinal data involving
The focus on means implies that all individuals eventually experience the final event, and it also requires a method to (in effect) impute the times of final events that are censored in the sample. As we noted in the Introduction, the former limitation can be relaxed by studying the age of occurrence of the final event within a restricted time window (10, 11). The method we applied to impute final event times for censored cases makes the common assumption that censoring is noninformative. This assumption could be relaxed by including auxiliary information in a model for imputing these times (28). The proposed measure can be readily extended to adjust for covariates, as when the effectiveness of a marker needs to be assessed net of other characteristics. If these characteristics are not time-varying, they can be included as covariates in the model used to estimate discriminatory ability.

Note that the measure of discriminatory ability at any age uses only the “current status” information of whether the marker has or has not occurred by that age. Thus, for cases where the marker has occurred, it does not use information about the time that has elapsed since marker occurrence. If this information were available, it could be used as a covariate to improve the prediction of time to the terminal event for individuals positive for the marker. This does not affect our aggregate measure of marker effectiveness, which averages over predictions of individuals with different elapsed times since marker occurrence.

ACKNOWLEDGMENTS

Author affiliations: Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan (Roderick J. Little, Bin Nan); and Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan (Matheos Yosef, Siobán D. Harlow).

This work was supported by grant AG 021543 (Siobán D. Harlow, Principal Investigator) from the National Institute on Aging.

The authors appreciate useful discussions on this work with Drs. Kevin Cain and Michael Elliott.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health.

Conflict of interest: none declared.

REFERENCES