Assessment of Digit Preference in Self-reported Year at Menopause: Choice of an Appropriate Reference Distribution

Sybil L. Crawford¹, Catherine B. Johannes², and Rebecca K. Stellato³

¹ Department of Medicine, Division of Preventive and Behavioral Medicine, Epidemiology and Biostatistics Unit, University of Massachusetts Medical Center, Worcester, MA.
² Epidemiology Division, Ingenix Pharmaceutical Services, Newton Lower Falls, MA.
³ Laboratory of Exposure Assessment and Environmental Epidemiology, National Institute of Public Health and the Environment, Bilthoven, the Netherlands.

Received for publication November 14, 2001; accepted for publication May 7, 2002.

Estimated associations between self-reported age at menopause and health may be attenuated if self-report is inaccurate. In a cross-sectional study, it is not possible to assess validity or reproducibility. Instead, one can examine digit preference, e.g., overreporting of numbers ending in zero or five. Typically, analyses use equal proportions—10% probability for each digit—as the reference distribution. Depending on the age distribution of the sample and on the underlying distribution of age at the event, however, an alternative reference distribution may be more appropriate. As an illustration, the authors examined digit preference in the self-reported calendar year at the final menstrual period in cross-sectional data from 2,151 naturally postmenopausal women in the Study of Women’s Health Across the Nation (1995–1997), a multisite, multiethnic study of women aged 40–55 years. With chi-square tests, the distribution of terminal digit for self-reported year at the final menstrual period was compared with several reference distributions. The observed distribution was much closer to a reference distribution based on previously published data than it was to equal proportions. Future assessments of digit preference in cross-sectional studies should consider alternatives to equal proportions, particularly for samples with small age ranges and events with a restricted underlying age distribution.

age of onset; menopause; recall

Abbreviations: NIA, National Institute on Aging; SWAN, Study of Women’s Health Across the Nation.

The timing of menopause has been linked to a number of health conditions, including cancer (1, 2), heart disease (3–6), osteoporosis and bone loss (7–11), and depression (12), as well as overall mortality (5, 13). Accurate measurement of age at menopause is important to avoid bias in these estimated associations (2, 14, 15). For example, inaccurate reporting has been found to attenuate the estimated risk of breast cancer with increasing age at menopause (2). Similarly, for diseases that may be negatively associated with age at menopause, such as cardiovascular disease or osteoporosis, the reduction of risk with a later menopause may be underestimated (15). Prospective measurement of age at menopause is most accurate (16, 17), but it is also the most expensive and logistically difficult study design. Note that the median time elapsed between the onset of perimenopause and the final menstrual period has been estimated as approximately 4 years (18, 19). A cross-sectional study can provide a timely, cost-efficient means of examining menopause and health (16, 20).

Analyses of cross-sectional data typically rely on subjects’ self-reported age or calendar year at menopause. The accuracy of this retrospectively recalled information has been assessed previously, primarily among Caucasians. Studies of validity have compared self-reports with medical records or prospectively collected data (15, 21–23), and studies of reproducibility have compared subjects’ self-reports from two separate interviews (15, 23–25). Both types of comparisons, however, require collection of information beyond that required for self-report of age at menopause.
Digit preference has been investigated in other health-related contexts, including blood pressure measurements (28–30), self-reported height and weight (31), and self-reported number of cigarettes smoked (32). Assessments of digit preference typically use as the reference distribution equal proportions, that is, 10 percent of the measurements are expected to end in each of the 10 digits zero through nine. Past studies of digit preference in recalled age at the final menstrual period have also used the equal proportions reference distribution, under the assumption that this represents unbiased reporting. Depending on the age distribution observed in the sample, however, an alternative reference distribution may be more appropriate. As an extreme example, consider studying only postmenopausal women aged 40–42 years. The resulting distribution of age at menopause would be highly restricted, and thus not all terminal digits would be equally likely.

The authors compared results from several reference distributions for assessing digit preference in self-reported data in the context of analyses of data from a cross-sectional interview of middle-aged women. The variable of interest was self-reported calendar year at the final menstrual period. In addition to equal proportions, alternative reference distributions were defined based on available data from internal and external sources to account for the relatively narrow age range in which the final menstrual period occurs and the age range of the sample.

**MATERIALS AND METHODS**

**Subjects**

Data were from the cross-sectional survey from the Study of Women’s Health Across the Nation (SWAN), a multisite, multiethnic, community-based study of health in middle-aged women. The survey, conducted in 1995 through 1997, involved a 15-minute telephone or in-person interview in seven geographic sites in the United States: Boston, Massachusetts; Chicago, Illinois; Detroit, Michigan; Los Angeles, California; Newark, New Jersey; Oakland, California; and Pittsburgh, Pennsylvania. Each site studied non-Hispanic, Caucasian women as well as women from one of the following racial/ethnic groups: African American (four sites), Chinese (one site), Hispanic (one site), and Japanese (one site). Race/ethnicity was determined based on the respondent’s primary self-identification. Eligibility criteria included residence in a designated geographic area, fluency in a designated language (English, Cantonese, Japanese, or Spanish), age between 40 and 55 years at the last birthday, and ability to provide verbal informed consent. The study was approved by each participating institution’s review board, and all subjects gave informed consent to participate. The study design has been described more fully elsewhere (33).

The SWAN cross-sectional interview had two goals: 1) to identify women eligible to participate in its longitudinal phase; and 2) to gather cross-sectional information on menopause and health in middle-aged women (33). Only approximately 1 percent of women reach natural menopause by age 40 years (34–37), and data from Caucasians with a natural menopause suggest that more than 80 percent reach postmenopause by age 55 years (18, 38, 39). Consequently, an age range of 40–55 years was selected to include the full set of ages at which women typically reach natural menopause.

Menopause status was determined from responses to questions regarding pregnancy, gynecologic surgery, and menstrual bleeding. Women who responded that they were not currently pregnant were asked whether they had had a hysterectomy and, if so, in what year. Nonpregnant women were also asked whether they had had both ovaries removed and, if so, in what year. Those who reported neither a hysterectomy nor a bilateral oophorectomy were asked whether they had had a menstrual period in the previous 12 months. Those who answered no were asked in what year their periods stopped and whether periods had stopped because of medication, chemotherapy, radiation treatment, pregnancy/breastfeeding, severe weight loss, or another clear reason except menopause. Women with 12 or more consecutive months of amenorrhea with no cause other than menopause were defined as naturally postmenopausal, following the standard definition (40–42).

Analyses presented here included the subgroup of women with a prevalent natural menopause and a nonmissing year of the final menstrual period. As noted above, the SWAN instruments asked subjects to recall the calendar year of the final menstrual period rather than their age at this event. Hence, analyses of terminal-digit preference were conducted for self-reported calendar year at menopause rather than for self-reported age at menopause. The methods used here also apply to analyses of terminal digit of age at an event.

**Statistical analyses**

The authors compared the distribution of the terminal digit of year of the final menstrual period with several reference distributions by using separate chi-square tests as follows. First, the reference distribution used was equal proportions. Second, an expected distribution for the terminal digit was estimated based on previously published data on the prevalence of natural menopause by current age in Caucasian women aged 45–55 years from the Massachusetts Women’s Health Study (38). Subjects in this study were asked whether they had had a period in the previous 12 months, and those who answered no and who had no other medical cause for amenorrhea were defined as naturally menopausal. Note that subjects were not asked for the year of or age at the final menstrual period, thereby minimizing inaccuracies from retrospective data collection. These data, omitting subjects with a hysterectomy or bilateral oophorectomy, were used to estimate a logistic regression model for the probability of prevalent natural menopause as a function of current age. From these estimated model coefficients, the authors esti-
mated the probability that the final menstrual period occurred at a particular age given that it had already occurred by the current age. For example, they computed the probability of the final menstrual period at age 45 years among postmenopausal women currently aged 50. After translation of age into calendar year to be consistent with the SWAN data, these estimated probabilities were applied to the observed distribution of age for each interview year in SWAN to produce an expected distribution of year at the final menstrual period and the resulting expected distribution of terminal digit of year at the final menstrual period. Further computational details are provided in the Appendix. Third, an analogous reference distribution was produced using estimated, age-specific prevalences of natural postmenopause from the SWAN cross-sectional data.

Fourth, observed interview year-specific distributions of the terminal digit were compared under the following rationale: If self-report is subject to digit preference, one would expect a similar distribution of the terminal digit for all interview years. If self-report is accurate, however, one would expect the distribution to be shifted systematically by one digit when comparing 2 consecutive years because the interview year changes by one digit. For example, under accurate self-reporting, the proportion with a terminal digit of one in 1997 should be similar to the proportion with a terminal digit of zero in 1996 and to the proportion with a terminal digit of nine in 1995. Consequently, the authors made two comparisons of terminal-digit distributions across interview years 1996 and 1997, omitting 1995 interviews because of a small sample size (n = 38). First, the original interview year-specific terminal-digit distributions were compared by using a chi-square test. A small chi-square statistic would be consistent with digit preference, whereas a large chi-square statistic would indicate differences by interview year. Second, the authors compared the distribution of the terminal digit shifted upward by one for 1996 subjects and the distribution of the terminal digit for 1997 subjects by using a chi-square test. A small chi-square statistic would be consistent with accurate reporting, while a large chi-square statistic would provide evidence of inaccurate self-reporting.

RESULTS

Among the SWAN cross-sectional sample of 15,273 women aged 40–55 years who self-identified with one of the five target ethnic groups, a total of 2,221 (14.5 percent) were naturally postmenopausal. Of these, 2,151 (96.8 percent) had a nonmissing year at the final menstrual period. Women in this subgroup were older and more likely to be Hispanic than were women who were premenopausal, surgically menopausal, or using hormone replacement therapy (table 1).

Comparison of terminal-digit distribution with equal proportions

Table 2 compares the observed distribution of the terminal digit with the equal-proportions reference distribution. A relatively high percentage of women (22.9 percent) had a terminal digit of four. The digit five also occurred more frequently than predicted under equal proportions (17.9 percent). The difference between the observed distribution and the equal-proportions reference distribution was highly significant. Note that the large observed percentages for the digit four and the low observed percentages for the digit zero were not consistent with the hypothesized peaks at digits zero and five.

Comparison of terminal-digit distribution with estimated reference distributions

Table 2 also presents chi-square statistics for the comparison of the observed terminal-digit distribution with the estimated reference distribution from the Massachusetts Women’s Health Study. The observed distributions were much closer to this reference distribution than to equal proportions, with a 13-fold reduction in the chi-square goodness-of-fit statistic. Figure 1 presents a graphic comparison of the observed and estimated terminal-digit distributions. Both the estimated reference distribution and the observed distribution exhibited a sharp drop at digit six as well as a rise from digits zero to three to a peak at digits four and five, although some remaining lack of fit was seen for digits four and five. The result for the SWAN-derived reference distribution was similar because of the comparability of the logistic regression models for the probability of prevalent postmenopause by age from the two datasets.

Terminal-digit distribution by year of interview

Table 3 presents the observed distributions of the terminal digit, stratifying by year of interview (1996–1997). The 1996 distribution appeared to be shifted downward by one digit relative to the 1997 distribution. This discrepancy was not accounted for by differences in age, as the age distribution did not vary significantly by interview year (p = 0.33). As
seen in figure 2, shifting the 1996 terminal digit by one reduced the overall difference between the 1996 and 1997 distributions. Both the original and the shifted terminal-digit distributions differed significantly by interview year, but the chi-square statistic for the latter was smaller than the former by a factor of 8.6. Thus, the interview year-specific results were more consistent with accurate self-report than with digit preference.

TABLE 2. Observed and estimated relative frequency distributions of the terminal digit for year at final menstrual period for 2,151 women with a prevalent natural menopause and a nonmissing year at the final menstrual period, Study of Women's Health Across the Nation, 1995–1997

<table>
<thead>
<tr>
<th>Terminal digit</th>
<th>Observed</th>
<th>Equal proportions</th>
<th>Massachusetts Women’s Health Study</th>
<th>SWAN†</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.4</td>
<td>10.0</td>
<td>7.0</td>
<td>6.2</td>
</tr>
<tr>
<td>1</td>
<td>8.0</td>
<td>10.0</td>
<td>9.4</td>
<td>8.7</td>
</tr>
<tr>
<td>2</td>
<td>11.7</td>
<td>10.0</td>
<td>12.0</td>
<td>11.6</td>
</tr>
<tr>
<td>3</td>
<td>13.5</td>
<td>10.0</td>
<td>14.8</td>
<td>15.0</td>
</tr>
<tr>
<td>4</td>
<td>22.9</td>
<td>10.0</td>
<td>17.5</td>
<td>18.4</td>
</tr>
<tr>
<td>5</td>
<td>17.9</td>
<td>10.0</td>
<td>19.8</td>
<td>21.5</td>
</tr>
<tr>
<td>6</td>
<td>7.2</td>
<td>10.0</td>
<td>8.8</td>
<td>9.4</td>
</tr>
<tr>
<td>7</td>
<td>2.3</td>
<td>10.0</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td>8</td>
<td>3.8</td>
<td>10.0</td>
<td>3.4</td>
<td>2.9</td>
</tr>
<tr>
<td>9</td>
<td>4.2</td>
<td>10.0</td>
<td>5.0</td>
<td>4.3</td>
</tr>
</tbody>
</table>

χ² statistic comparing observed with reference

<table>
<thead>
<tr>
<th>Terminal digit</th>
<th>Observed</th>
<th>Equal proportions</th>
<th>Massachusetts Women’s Health Study</th>
<th>SWAN†</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.4</td>
<td>10.0</td>
<td>7.0</td>
<td>6.2</td>
</tr>
<tr>
<td>1</td>
<td>8.0</td>
<td>10.0</td>
<td>9.4</td>
<td>8.7</td>
</tr>
<tr>
<td>2</td>
<td>11.7</td>
<td>10.0</td>
<td>12.0</td>
<td>11.6</td>
</tr>
<tr>
<td>3</td>
<td>13.5</td>
<td>10.0</td>
<td>14.8</td>
<td>15.0</td>
</tr>
<tr>
<td>4</td>
<td>22.9</td>
<td>10.0</td>
<td>17.5</td>
<td>18.4</td>
</tr>
<tr>
<td>5</td>
<td>17.9</td>
<td>10.0</td>
<td>19.8</td>
<td>21.5</td>
</tr>
<tr>
<td>6</td>
<td>7.2</td>
<td>10.0</td>
<td>8.8</td>
<td>9.4</td>
</tr>
<tr>
<td>7</td>
<td>2.3</td>
<td>10.0</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td>8</td>
<td>3.8</td>
<td>10.0</td>
<td>3.4</td>
<td>2.9</td>
</tr>
<tr>
<td>9</td>
<td>4.2</td>
<td>10.0</td>
<td>5.0</td>
<td>4.3</td>
</tr>
</tbody>
</table>

χ² statistic comparing observed with reference

<table>
<thead>
<tr>
<th>Terminal digit</th>
<th>Observed</th>
<th>Equal proportions</th>
<th>Massachusetts Women’s Health Study</th>
<th>SWAN†</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8.4</td>
<td>10.0</td>
<td>7.0</td>
<td>6.2</td>
</tr>
<tr>
<td>1</td>
<td>8.0</td>
<td>10.0</td>
<td>9.4</td>
<td>8.7</td>
</tr>
<tr>
<td>2</td>
<td>11.7</td>
<td>10.0</td>
<td>12.0</td>
<td>11.6</td>
</tr>
<tr>
<td>3</td>
<td>13.5</td>
<td>10.0</td>
<td>14.8</td>
<td>15.0</td>
</tr>
<tr>
<td>4</td>
<td>22.9</td>
<td>10.0</td>
<td>17.5</td>
<td>18.4</td>
</tr>
<tr>
<td>5</td>
<td>17.9</td>
<td>10.0</td>
<td>19.8</td>
<td>21.5</td>
</tr>
<tr>
<td>6</td>
<td>7.2</td>
<td>10.0</td>
<td>8.8</td>
<td>9.4</td>
</tr>
<tr>
<td>7</td>
<td>2.3</td>
<td>10.0</td>
<td>2.3</td>
<td>1.9</td>
</tr>
<tr>
<td>8</td>
<td>3.8</td>
<td>10.0</td>
<td>3.4</td>
<td>2.9</td>
</tr>
<tr>
<td>9</td>
<td>4.2</td>
<td>10.0</td>
<td>5.0</td>
<td>4.3</td>
</tr>
</tbody>
</table>

χ² statistic comparing observed with reference

*p < 0.001.
† SWAN, Study of Women’s Health Across the Nation.

FIGURE 1. Distribution of the terminal digit for year of the final menstrual period for 2,151 women with a prevalent natural menopause and nonmissing year at the final menstrual period, Study of Women’s Health Across the Nation, 1995–1997.
DISCUSSION

In these cross-sectional data on self-reported calendar year at menopause, the distribution of the terminal digit differed sharply from equal proportions. This divergence, however, does not necessarily reflect inaccurate self-reporting. Note that the peaks at digits zero and five, hypothesized under an equal-proportions reference, did not occur. In addition, the observed distribution was similar to expected distributions based on previously published prevalence data and on prevalence data from SWAN. The comparison of interview year-specific distributions also suggested that digit preference was not operating. On the basis of the observed lack of digit preference, analyses of cross-sectional survey data from SWAN regarding the relation between age at menopause and health conditions, such as heart disease (43), should not be subject to bias due to inaccurately reported age at menopause.

The equal proportions reference distribution is appropriate for assessing digit preference in many situations, including blood pressure measurement and self-reported height and weight. In such cases, the underlying distributions often can be expected to be close to a normal distribution over a wide range of possible values. In the example presented here, however, the expected distribution of the terminal digit was concentrated on several consecutive digits rather than spread equally over all 10 digits because of the relatively narrow ranges of both the age at the event of interest and participant age. Similar issues may arise for other health-related events of interest with a restricted age distribution, such as age at menarche (44–47), gestational age in fetal births and deaths (48–50), age at onset of psychiatric symptoms in children (51), and age at first sexual intercourse (52).

As noted above, inaccurately recalled, self-reported age or year at onset can yield estimates of associations with diseases that are biased toward the null (no association).

---

### TABLE 3. Observed relative frequency distribution of the terminal digit for year at the final menstrual period for 2,113 women with a prevalent natural menopause and a nonmissing year at the final menstrual period, stratified by year of interview, Study of Women’s Health Across the Nation, 1996–1997*

<table>
<thead>
<tr>
<th>Terminal digit</th>
<th>% with terminal digit for year of final menstrual period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1996 interview subjects (n = 1,335)</td>
</tr>
<tr>
<td>0</td>
<td>9.1</td>
</tr>
<tr>
<td>1</td>
<td>9.8</td>
</tr>
<tr>
<td>2</td>
<td>11.5</td>
</tr>
<tr>
<td>3</td>
<td>15.0</td>
</tr>
<tr>
<td>4</td>
<td>27.3</td>
</tr>
<tr>
<td>5</td>
<td>14.1</td>
</tr>
<tr>
<td>6</td>
<td>2.3</td>
</tr>
<tr>
<td>7</td>
<td>1.6</td>
</tr>
<tr>
<td>8</td>
<td>4.2</td>
</tr>
<tr>
<td>9</td>
<td>4.9</td>
</tr>
</tbody>
</table>

* $\chi^2$ statistic for independence of original interview year-specific distributions = 226.91 ($p < 0.001$). $\chi^2$ statistic for independence of shifted interview year-specific distributions = 26.54 ($p < 0.001$).

---

**FIGURE 2.** Distribution of the terminal digit for year of the final menstrual period for 2,113 women with a prevalent natural menopause and nonmissing year at the final menstrual period, Study of Women’s Health Across the Nation, 1996–1997.
Results regarding accuracy of self-report from past studies may not always apply, particularly when investigating populations not previously well-studied, such as non-Caucasians in SWAN. Thus, to the extent possible, it is important to assess directly the accuracy of this information prior to drawing conclusions from one’s own data. In many studies, particularly those with cross-sectional designs, reproducibility and validity cannot be determined (25). Instead, the available information, both internal and external to an investigator’s study, can be analyzed with techniques similar to those used here. We have shown that the use of the equal-proportions reference distribution is inadequate for assessing digit preference with the type of data considered here, that is, involving a restricted age range. If available, reference distributions from other studies with similar data can provide a useful and accurate way to assess digit preference. In addition, a simple and cost-efficient technique, stratification by interview year, is another way to gauge the accuracy of self-reported age or calendar year at onset.

ACKNOWLEDGMENTS

SWAN was funded by the National Institute on Aging (NIA), the National Institute of Nursing Research, and the Office of Research on Women’s Health of the National Institutes of Health. Supplemental funding from the National Institute of Mental Health, the National Institute on Child and Human Development, the National Center on Complementary and Alternative Medicine, and the Office of AIDS Research is also gratefully acknowledged.

Clinical centers: University of Michigan, Ann Arbor, MI—Dr. MaryFran Sowers, Principal Investigator (National Institute of Nursing Research grant U01 NR04061); Massachusetts General Hospital, Boston, MA—Dr. Robert Neer, Principal Investigator, 1995–1999, Dr. Joel Finkelstein, Principal Investigator, 1999–present (NIA grant U01 AG12531); Rush University, Rush-Presbyterian-St. Luke’s Medical Center, Chicago, IL—Dr. Lynda Powell, Principal Investigator (NIA grant U01 AG12505); University of California, Davis/Kaiser, Davis, CA—Dr. Ellen Gold, Principal Investigator (NIA grant U01 AG12554); University of California, Los Angeles, CA—Dr. Gail Greendale, Principal Investigator (NIA grant U01 AG12539); University of Medicine and Dentistry-New Jersey Medical School, Newark, NJ—Dr. Gerson Weiss, Principal Investigator (NIA grant U01 AG12535); and the University of Pittsburgh, Pittsburgh, PA—Dr. Karen Matthews, Principal Investigator (NIA grant U01 AG12546).

National Institutes of Health Project Office: National Institute on Aging, Bethesda, MD—Drs. Sheryl Sherman and Marcia Ory; National Institute of Nursing Research, Bethesda, MD—Dr. Carole Hudgings.

Central Laboratory: University of Michigan, Ann Arbor, MI—Dr. Rees Midgley, Principal Investigator, 1995–2000, Dr. Dan McConnell, Principal Investigator, 2000–present (NIA grant U01 AG12495, Central Ligand Assay Satellite Services).

Coordinating Center: New England Research Institutes, Watertown, MA—Dr. Sonja McKinlay, Principal Investigator, 1995–2001; the University of Pittsburgh, Pittsburgh, PA—Dr. Kim Sutton-Tyrrell, Principal Investigator, 2001–present (NIA grant U01 AG12553).

Steering Committee: Dr. Christopher Gallagher, Chair, 1995–1997; Dr. Jennifer Kelsey, Chair, 1997–present.

The authors thank the study staff at each site. The authors also gratefully acknowledge the advice of Professor Alan Agresti, University of Florida, regarding exact testing.

The authors were affiliated with New England Research Institutes, Watertown, MA, during much of the work on this paper. The manuscript was reviewed by the Publications and Presentations Committee of SWAN and has its endorsement.

REFERENCES


APPENDIX

The expected reference distribution based on the observed interview year-specific age distributions of SWAN and on previously published data from the Massachusetts Women’s Health Study (38) was computed as follows. Denote the year of the final menstrual period by FMPYR, the subject’s age at the final menstrual period by FMPAGE, the subject’s age at the time of the SWAN interview by t, and the observed age distribution of the full sample of nonsurgical nonusers of hormone replacement therapy at interview t by Zt.
Step 1
The probability that $FMPAGE = j$ among naturally postmenopausal women at the $t$th interview is given by:

$$
\Pr(FMPAGE = j | \text{post at interview } t, Zt) = \sum_{i=40}^{55} \Pr(FMPAGE = j | A = i, \text{post at interview } t, Zt) \times \Pr(A = i | \text{post at interview } t, Zt).
$$

Step 2
The first term in the summation above is calculated as:

$$
\Pr(FMPAGE = j | A = i, \text{post at interview } t, Zt) = \Pr(FMPAGE \leq j \leq i - 1, Zt) / \Pr(FMPAGE \leq i - 1 | Zt) \text{ for } j \leq i - 1, 0 \text{ for } j > i - 1.
$$

McKinlay et al. (38) present age-specific percentages of naturally postmenopausal women among nonsurgically menopausal women not using hormone replacement therapy. Note that $\Pr(FMPAGE \leq c | Zt)$ can be estimated as $\exp(\logit)/(1 + \exp(\logit))$, where $\logit = \beta_0 + \beta_1 \times c$, and $\beta_0$ and $\beta_1$ are from the logistic regression model estimated from these prevalence data, regressing the probability of a prevalent natural menopause on current age.

Step 3
The second term in the summation in step 1 is calculated as:

$$
\Pr(A = i | \text{post at interview } t, Zt) = \Pr(\text{post at interview } t | A = i, Zt) \times \Pr(A = i | Zt) / \Pr(\text{post at interview } t | Zt).
$$

Each of these terms can be computed as described above.

Step 4
The expected distribution for $FMPAGE$ is translated into the corresponding expected distribution for $FMPYR$ by using the following relation: $FMPYR = FMPAGE - (A - \text{interview year})$.

Step 5
The expected distribution for $FMPYR$ over all 3 interview years is a weighted average of the expected interview year-specific distributions, weighting by the relative proportion of women expected to be postmenopausal given the observed age distribution in that interview year.

Step 6
The expected distribution for the terminal digit of $FMPYR$ is computed by summing probabilities for $FMPYR$ corresponding to the terminal digit, that is, the probability that the terminal digit is zero is computed by summing probabilities for $FMPYR = 10, 20, 30, 40,$ and so forth.