ESTIMATION OF THE DISTRIBUTION OF AGE AT NATURAL MENOPAUSE FROM PREVALENCE DATA

M. D. KRAILO AND M. C. PIKE


Nearly 30% of US women reach menopause (defined as cessation of menstrual periods) as a consequence of an operation. This biases the observable distribution of age at natural menopause. Another problem with estimating this distribution from a cross-sectional study is the clustering of reported age at natural menopause around ages ending in zero and five (MacMahon B, Worcester J. Age at menopause, United States 1960–1962. Washington DC: National Center for Health Statistics, 1966. Vital and health statistics, Series 11: Data from the National Health Survey, no. 19. (DHEW publication no. (HSM) 66-1000)). This paper discusses the approach of MacMahon and Worcester to this problem and compares it with a competing risks approach.

In the National Center for Health Statistics' Health Examination Survey (1), conducted over the years 1960–1962, data were obtained on the menopausal history of 3581 "representative" US women. The relevant questions asked were:

(a) Have your periods stopped? Yes/No
(b) If Yes to (a)
   (b.1) Age when periods stopped (years)
   (b.2) Was this due to an operation? Yes/No

On the basis of answers to these questions and the age of the women at the time of response to the survey, it is possible, using standard life-table methods, to calculate the "observed" distribution of age at natural menopause (defined as non-operative cessation of menstrual periods) simply regarding operative menopause as a "competing risk." This observed distribution may also be fitted to a hypothesized parametric distribution using standard methods.

However, when MacMahon and Worcester (1) looked closely at this important data set, they found that there was "marked terminal digit clustering" of reported age at natural menopause around ages ending in zero or five, and they concluded that standard life-table methods would give erroneous results. They therefore decided to ignore the answers to question b.1 and to estimate the distribution of age at natural menopause using only the quantal (Yes/No) answers. The data they considered, covering 2423 women, are given in table 1 as "observed data."

For reasons that are not completely clear, but certainly including ease of computation, MacMahon and Worcester (1) also decided in their analysis of age at natural menopause to exclude all women who had had an operative menopause, even though slightly more than one quar-
<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Age midpoint $t$</th>
<th>Responders $R$</th>
<th>Operative menopause $X$</th>
<th>Natural menopause $Y$</th>
<th>Still menstruating $Z$</th>
<th>Operative menopause</th>
<th>Natural menopause</th>
<th>Still menstruating</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-29</td>
<td>27.5</td>
<td>380</td>
<td>4</td>
<td>0</td>
<td>376</td>
<td>4.81</td>
<td>0</td>
<td>375.19</td>
</tr>
<tr>
<td>30-34</td>
<td>32.5</td>
<td>359</td>
<td>21</td>
<td>0</td>
<td>338</td>
<td>16.27</td>
<td>0</td>
<td>342.73</td>
</tr>
<tr>
<td>35</td>
<td>35.5</td>
<td>89</td>
<td>7</td>
<td>0</td>
<td>82</td>
<td>6.57</td>
<td>0.04</td>
<td>82.39</td>
</tr>
<tr>
<td>36</td>
<td>36.5</td>
<td>87</td>
<td>5</td>
<td>0</td>
<td>82</td>
<td>7.36</td>
<td>0.14</td>
<td>79.50</td>
</tr>
<tr>
<td>37</td>
<td>37.5</td>
<td>61</td>
<td>5</td>
<td>1</td>
<td>55</td>
<td>5.86</td>
<td>0.20</td>
<td>54.93</td>
</tr>
<tr>
<td>38</td>
<td>38.5</td>
<td>83</td>
<td>11</td>
<td>2</td>
<td>70</td>
<td>8.97</td>
<td>0.49</td>
<td>73.53</td>
</tr>
<tr>
<td>39</td>
<td>39.5</td>
<td>99</td>
<td>11</td>
<td>2</td>
<td>86</td>
<td>11.95</td>
<td>0.97</td>
<td>86.08</td>
</tr>
<tr>
<td>40</td>
<td>40.5</td>
<td>78</td>
<td>8</td>
<td>1</td>
<td>69</td>
<td>10.43</td>
<td>1.21</td>
<td>66.36</td>
</tr>
<tr>
<td>41</td>
<td>41.5</td>
<td>66</td>
<td>7</td>
<td>1</td>
<td>58</td>
<td>9.71</td>
<td>1.57</td>
<td>54.72</td>
</tr>
<tr>
<td>42</td>
<td>42.5</td>
<td>80</td>
<td>16</td>
<td>4</td>
<td>60</td>
<td>12.87</td>
<td>2.87</td>
<td>64.26</td>
</tr>
<tr>
<td>43</td>
<td>43.5</td>
<td>74</td>
<td>11</td>
<td>5</td>
<td>58</td>
<td>12.93</td>
<td>3.94</td>
<td>57.13</td>
</tr>
<tr>
<td>44</td>
<td>44.5</td>
<td>67</td>
<td>10</td>
<td>3</td>
<td>54</td>
<td>12.65</td>
<td>5.20</td>
<td>49.16</td>
</tr>
<tr>
<td>45</td>
<td>45.5</td>
<td>99</td>
<td>20</td>
<td>12</td>
<td>67</td>
<td>20.05</td>
<td>10.99</td>
<td>67.96</td>
</tr>
<tr>
<td>46</td>
<td>46.5</td>
<td>76</td>
<td>16</td>
<td>11</td>
<td>49</td>
<td>16.40</td>
<td>11.80</td>
<td>47.80</td>
</tr>
<tr>
<td>47</td>
<td>47.5</td>
<td>75</td>
<td>18</td>
<td>16</td>
<td>41</td>
<td>17.12</td>
<td>15.83</td>
<td>42.04</td>
</tr>
<tr>
<td>48</td>
<td>48.5</td>
<td>80</td>
<td>19</td>
<td>18</td>
<td>43</td>
<td>19.18</td>
<td>22.20</td>
<td>38.62</td>
</tr>
<tr>
<td>49</td>
<td>49.5</td>
<td>66</td>
<td>20</td>
<td>19</td>
<td>27</td>
<td>16.48</td>
<td>23.14</td>
<td>26.38</td>
</tr>
<tr>
<td>50</td>
<td>50.5</td>
<td>72</td>
<td>18</td>
<td>32</td>
<td>22</td>
<td>18.59</td>
<td>30.59</td>
<td>22.82</td>
</tr>
<tr>
<td>51</td>
<td>51.5</td>
<td>66</td>
<td>10</td>
<td>38</td>
<td>18</td>
<td>17.49</td>
<td>32.63</td>
<td>15.89</td>
</tr>
<tr>
<td>52</td>
<td>52.5</td>
<td>54</td>
<td>16</td>
<td>30</td>
<td>8</td>
<td>14.59</td>
<td>29.92</td>
<td>9.49</td>
</tr>
<tr>
<td>53</td>
<td>53.5</td>
<td>67</td>
<td>18</td>
<td>40</td>
<td>9</td>
<td>18.36</td>
<td>40.33</td>
<td>8.31</td>
</tr>
<tr>
<td>54</td>
<td>54.5</td>
<td>50</td>
<td>18</td>
<td>28</td>
<td>4</td>
<td>13.84</td>
<td>31.90</td>
<td>4.26</td>
</tr>
<tr>
<td>55</td>
<td>55.5</td>
<td>45</td>
<td>19</td>
<td>28</td>
<td>1</td>
<td>12.55</td>
<td>29.87</td>
<td>2.58</td>
</tr>
<tr>
<td>56</td>
<td>56.5</td>
<td>50</td>
<td>13</td>
<td>36</td>
<td>1</td>
<td>14.01</td>
<td>34.09</td>
<td>1.90</td>
</tr>
<tr>
<td>57</td>
<td>57.5</td>
<td>54</td>
<td>13</td>
<td>40</td>
<td>1</td>
<td>15.18</td>
<td>37.48</td>
<td>1.35</td>
</tr>
<tr>
<td>58</td>
<td>58.5</td>
<td>46</td>
<td>13</td>
<td>33</td>
<td>0</td>
<td>12.96</td>
<td>32.30</td>
<td>0.75</td>
</tr>
</tbody>
</table>

* Chi-square statistic: 22.65 on 36 degrees of freedom, $p > 0.50$. 
ter of the women surveyed would reach menopause in this way. The authors were aware that eliminating women in the operative menopause category would lead to a biased estimate of the distribution of age at natural menopause, “since increases in the incidence of operative menopause from one age to another will produce an apparent increase in the prevalence of natural menopause as it has been defined” (1, p. 7).

In this paper we develop a competing risks model for age at menopause, describe how it may be applied to prevalence data, show that it provides an excellent fit to the data of table 1, and finally discuss the accuracy of MacMahon and Worcester’s simplifying assumption on the basis of comparisons with this competing risks model.

**Materials and Methods**

**Menopausal incidence rates**

We assume a standard competing risks model (2) with two risks: natural menopause and operative menopause. We furthermore assume that intercurrent mortality is independent of these risks and thus that it can be ignored when dealing with prevalence data.

**Natural menopause.** Following MacMahon and Worcester (1), we assume that the “underlying” distribution of age at natural menopause is logistic, so that the incidence rate (hazard rate), \( n(t) \), at age \( t \) (years) is

\[
n(t) = b \exp(a + bt)/[1 + \exp(a + bt)]
\]

for some constants \( a \) and \( b \).

**Operative menopause.** MacMahon and Worcester did not find terminal digit clustering of stated age at operative menopause, and made crude estimates of the conditional probabilities of operative menopause among women still menstruating (1, table A). These estimates show that, as a first approximation, the incidence rate, \( h(t) \), of operative menopause simply increases linearly from about age 22 years, i.e., we suppose that

\[
h(t) = c(t - 22)
\]

for some constant \( c \) for \( t \geq 22 \) and is zero for \( t < 22 \).

The adequacy of fit of equations 1 and 2 to the data of table 1 is discussed below. The incidence rate, \( m(t) \), for either natural or operative menopause is therefore:

\[
m(t) = n(t) + h(t)
\]

(see reference 2, p. 167).

**Construction of the likelihood function**

The \( i \)th line of data (for \( i = 1, 2, \ldots, k \)) in table 1 may be written \((t_i, R_i, X_i, Y_i)\), where: \( t_i = \) age (years) at response (set at the midpoint of the age interval); \( R_i = \) number of women responding at age \( t_i \); \( X_i = \) number of these \( R_i \) women having had operative menopause; and \( Y_i = \) number of these \( R_i \) women having had natural menopause.

The number, \( Z_i \), of these \( R_i \) women who were still menstruating is given by

\[
Z_i = R_i - X_i - Y_i
\]

The following three probabilities are needed to write down the likelihood for this \( i \)th line of data:

\[
F(t_i) = \text{Prob}(\text{no menopause by age } t_i),
\]

\[
H(t_i) = \text{Prob}(\text{operative menopause by age } t_i)
\]

(i.e., probability periods ceased due to operation by age \( t_i \)), and

\[
N(t_i) = \text{Prob}(\text{natural menopause by age } t_i)
\]

(i.e., probability periods ceased due to natural menopause by age \( t_i \)).

The formulae for \( F(t) \), \( H(t) \), and \( N(t) \) in terms of \( n(t) \) and \( h(t) \), and hence in terms of the unknown parameters \( a \), \( b \), and \( c \), are given in the Appendix.
DISTRIBUTION OF AGE AT NATURAL MENOPAUSE

The likelihood for the $i$th line of data $(t_i, R_i, X_i, Y_i)$ is

$$l_i = [H(t_i)]^{x_i}[N(t_i)]^{y_i}[F(t_i)]^{z_i}$$  (7)

and the total likelihood for the complete data set is

$$l = \prod_{i=1}^{k} l_i.$$  (8)

The likelihood function was maximized by a direct search algorithm (3).

The goodness of fit of the model to the data was assessed in the following manner: the fitted frequencies $(\hat{X}_i, \hat{Y}_i)$, of women having achieved operative or natural menopause for each age group given in table 1, and the estimated variance-covariance matrix $\hat{V}_i$ of the frequencies, were calculated using the estimated parameters. Age groups were then collapsed to insure that for all groups, both $\hat{X}_i$ and $\hat{Y}_i$ were greater than or equal to three. The statistic $D = \sum (\hat{X}_i\hat{Y}_i)\hat{V}_i(\hat{X}_i\hat{Y}_i)'$ was used to assess the fit of the data. Shillington (4) has shown that this statistic is distributed as approximately chi-square with number of degrees of freedom equal to (twice the number of remaining age groups - number of estimated parameters - 1).

RESULTS

Maximization of the total likelihood function (equation 8) yielded estimates of the parameters: $\hat{a} = -20.7, \hat{b} = 0.414$ and $\hat{c} = 0.000841$; with estimated variance-covariance matrix:

$$2.23 \quad -0.0451 \quad 5.00 \times 10^{-5}$$
$$- \quad 0.000915 \quad -1.01 \times 10^{-6}$$
$$- \quad - \quad 3.05 \times 10^{-9}$$

calculated as the inverse of the observed Fisher information matrix. The fitted values for the numbers of individuals in the operative menopause and natural menopause groups are given in columns 6 and 7 of table 1, and graphs of the observed and fitted proportions of women achieving the two types of menopause at various ages are plotted against age in figure 1. The chi-square goodness of fit statistic and graph indicate an excellent fit (22.65 on 36 degrees of freedom, $p > 0.50$).

DISCUSSION

The fitted parameters for the "underlying" presumed logistic distribution of age at natural menopause are very close to those obtained by the method of MacMahon and Worcester (1). The present method gives $\hat{a} = -20.7$ and $\hat{b} = 0.414$; this compares to MacMahon and Worcester's $\hat{a} = -21.5$ and $\hat{b} = 0.43$.

The competing risks model described above allows accurate investigation of the distribution of age at natural menopause while accounting for the effects of operative menopause. If the incidence rate, $n(t)$, for natural menopause is "independent" of $h(t)$, i.e., if $n(t)$ is unchanged by changes in $h(t)$ (including its elimination—see reference 2, p. 178), then, in addition, this

\[ \text{FIGURE 1. Observed and fitted percentages of the study respondents having achieved operative (•) or natural (○) menopause as a function of age.} \]
Table 2

<table>
<thead>
<tr>
<th>$c$</th>
<th>Probability of natural menopause (%)</th>
<th>Average age (years) at natural menopause</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100.0</td>
<td>50.0</td>
</tr>
<tr>
<td>$\frac{c}{2}$</td>
<td>71.8</td>
<td>49.5</td>
</tr>
<tr>
<td>$2c$</td>
<td>52.0</td>
<td>49.2</td>
</tr>
</tbody>
</table>

* Assuming the incidence rate for natural menopause is fixed for all values of $c$ (see Discussion).

The above model does not incorporate any terms reflecting secular changes in operative menopause. This is because the data are cross-sectional and secular trends are confounded by age. Even though rates have fluctuated over time the maximal rates observed in the early 1970s are still only one and a half times the estimated rates given here (5), and the results of table 2 suggest that ignoring secular trends will not have produced a major error. Thus, comparisons of age at natural menopause between different populations made from prevalence studies are likely to be confounded to only a small degree by differential rates of operative menopause.

References

APPENDIX

With the hypothesized forms for the incidence rates given by equations 1, 2, and 3, the quantities given in expressions 4, 5, and 6 may be expressed as:

\[ F(t) = \exp\left[-\int_0^t m(x)dx\right], \]
\[ H(t) = \int_0^t h(x)F(x)dx \]
and
\[ N(t) = \int_0^t n(x)F(x)dx. \]

The following quantities are also of interest:

\[ P(\text{natural menopause}) = \int_0^\infty n(t)F(t)dt, \]
and
\[ E(\text{age at natural menopause}|\text{natural menopause occurs}) = \int_0^\infty t n(t)F(t)dt/P(\text{natural menopause}). \]

Maximum likelihood estimates of all these quantities are obtained by substituting the maximum likelihood estimates of \(a, b,\) and \(c\) as appropriate.