7 Appendix

Formulae for estimation of expected survival

Under the Ederer I method (Ederer et al. 1961), the cumulative expected survival from the date of diagnosis to the end of the \( i \)th interval is given by

\[
1p^*_{i} = \frac{\sum_{h=1}^{l_1} 1p^*_i (h)}{l_1},
\]

where \( l_1 \) is the total number of patients alive at the start of follow-up and \( 1p^*_i (h) \) is the expected probability of surviving to the end of the \( i \)th interval for a person in the general population, similar to the \( h \)th patient alive at the beginning of follow-up with respect to age, sex, and calendar time, given by

\[
1p^*_i (h) = \prod_{j=1}^{i} p^*_j (h).
\]

Under the Ederer II method (Ederer and Heise 1959)

\[
1p^*_i = \prod_{j=1}^{i} p^*_{j2},
\]

where

\[
p^*_{j2} = \frac{\sum_{h=1}^{l_j} p^*_j (h)}{l_j}
\]

is the average of the annual expected survival probabilities \( p^*_j (h) \) of the patients alive at the start of the \( j \)th interval.

The expected survival proportion using the Hakulinen method (Hakulinen 1982) is derived as follows. Let \( k_j \) be the number of patients with a potential follow-up time which extends beyond the beginning of the \( j \)th interval. Let the first \( k_{ja} \) of these \( k_j \) patients have a potential follow-up time which extends past the end of the \( j \)th interval and the last \( k_{jb} \) be potential withdrawals during the \( j \)th interval. It follows that \( k_1 = l_1 \), \( k_{j+1} = k_{ja} \), and \( k_j = k_{ja} + k_{jb} \). We will use the notation \( K_{ja} \) to refer to the set of \( k_{ja} \) patients etc. and \( h \) to index the \( k_{ja} \) patients in the set \( K_{ja} \). The expected number of patients alive and under observation at the beginning of the \( j \)th interval is given by:

\[
l^*_j = \begin{cases} 
\sum_{h \in K_j} 1p^*_{j-1} (h) & \text{for } j \geq 2 \\
l_1 & \text{for } j = 1
\end{cases}
\]

For the \( k_{jb} \) patients with potential follow-up times ending during the \( j \)th interval, it is assumed that each patient is at risk for half of the interval, so the expected probability of dying during the interval is given by \( 1 - \sqrt{p^*_j} \). The expected number of patients
withdrawing alive during the \( j \)th interval is therefore given by:

\[
    w_j^* = \begin{cases} 
    \sum_{h \in K_{jb}} p_{j-1}^*(h) \sqrt{p_j^*(h)} & \text{for } j \geq 2 \\
    \sum_{h \in K_{1b}} \sqrt{p_1^*(h)} & \text{for } j = 1 
    \end{cases}
\]

The expected number of patients dying during the \( j \)th interval, among the \( k_{jb} \) patients with potential follow-up time ending during the same interval is given by:

\[
    \delta_j^* = \begin{cases} 
    \sum_{h \in K_{jb}} p_{j-1}^*(h)[1 - \sqrt{p_j^*(h)}] & \text{for } j \geq 2 \\
    \sum_{h \in K_{1b}} [1 - \sqrt{p_1^*(h)}] & \text{for } j = 1 
    \end{cases}
\]

and the expected total number of patients dying during the \( j \)th interval is given by:

\[
    d_j^* = \begin{cases} 
    \left\{ \sum_{h \in K_{ja}} p_{j-1}^*(h)[1 - p_j^*(h)] \right\} + \delta_j^* & \text{for } j \geq 2 \\
    \left\{ \sum_{h \in K_{1a}} [1 - p_1^*(h)] \right\} + \delta_1^* & \text{for } j = 1 
    \end{cases}
\]

The expected interval-specific survival proportion is then written as:

\[
    g_j^* = 1 - d_j^*/(l_j^* - w_j^*/2),
\]

and, finally, the expected survival proportion from the beginning of follow-up (usually diagnosis) to the end of the \( i \)th interval is obtained by calculating:

\[
    \hat{p}_i^* = \prod_{j=1}^{i} g_j^*.
\]