Decomposition by tumor size, cancer-specific mortality rates, and competing causes of death

Let $\pi_i(t)$ and $e_i(x,t)$ be the proportion of patients and the life expectancy for cancer patients with tumor size i (i.e., $<$ 1 cm, 1-2 cm, 2-3 cm, 3-5 cm, and 5+ cm.) at age x and time t, respectively. That is, $e_i(x,t)$ represents tumor-size-specific life expectancy. The overall life expectancy at age x time t is given by

$$e(x,t) = \sum_{i=1}^{5} \pi_i(t)\, e_i(x,t)$$

where $\sum_{i=1}^{5} \pi_i = 1$. Consider tumor size i. The change in life expectancy at age x between times t_1 and t_2 can be decomposed as (Kitigawa 1955):

$$e_i(x,t_2) - e_i(x,t_1) = \sum_{j=1}^{k} \int_x^{\infty} \left[p_{i,j}(s,t_2) - p_{i,j}(s,t_1)\right] \left[\frac{p_{i,-j}(s,t_1) + p_{i,-j}(s,t_2)}{2}\right] ds,$$

where $p_{i,j}(s,t)$ is the probability of surviving from birth to age s for cancer patients with tumor size i cause j at time t, and $p_{i,-j}(s,t)$ is the analogue survival probability for all other causes of death (other than j). Similarly, the change in life expectancy across age tumor sizes can be decomposed as:

$$\begin{aligned} e(x,t_2) - e(x,t_1) &= \sum_{i=0}^{3} \left[\, \pi_i(t_2)\, e_i(x,t_2) - \pi_i(t_1)\, e_i(x,t_1)\right] \\ &= \sum_{i=0}^{3} \left[\pi_i(t_2) - \pi_i(t_1)\right] \left[\frac{e_i(x,t_1) + e_i(x,t_2)}{2}\right] + \sum_{i=0}^{3} \left[e_i(x,t_2) - e_i(x,t_1)\right] \left[\frac{\pi_i(t_1) + \pi_i(t_2)}{2}\right]. \end{aligned}$$

The above equation quantifies how much of the change in life expectancy at age x between times t_1 and t_2 is due to: [a] shifts in the share of cancer tumor size (first term) and [b] changes in tumor-size-specific life expectancy (second term).

We can further decompose the second term in the above equation by cause of death. In doing so, we can quantify how much of this change in cancer-specific life expectancy, $e_i(x,t_2)-e_i(x,t_1)$, is due to improvements in cancer mortality and non-cancer mortality. Using the approach developed in Beltrán-Sánchez et al. (2008),

$$e_i(x,t_2) - e_i(x,t_1) = \sum_{j=1}^{2} \sum_{s=x}^{\omega} \left[L_{s,i,j}(t_2) - L_{s,i,j}(t_1)\right] \left[\frac{L_{s,i,-j}(t_2) + L_{s,i,-j}(t_1)}{2n}\right] \tag{0.1}$$

where i corresponds to tumor size, j is cause-specific mortality among patients diagnosed with tumor size i, s is age, ω is the starting age of the oldest age interval, n is the width of the age interval, and L_s are person-years lived in the life table.

We perform the decomposition starting at age 40, so the final decomposition equation is given by:

$$\begin{aligned} e(40,t_2) - e(40,t_1) &= \sum_{i=1}^{5} \left[\, \pi_i(t_2)\, e_i(40,t_2) - \pi_i(t_1)\, e_i(40,t_1)\right] \\ &= \sum_{i=1}^{5} \left[\pi_i(t_2) - \pi_i(t_1)\right] \left[\frac{e_i(40,t_1) + e_i(40,t_2)}{2}\right] + \sum_{i=1}^{5} \left[\texttt{Diff}\right] \left[\frac{\pi_i(t_1) + \pi_i(t_2)}{2}\right] \end{aligned} \tag{0.2}$$

where `Diff` is given by (0.1) evaluated at x=40.

Decomposition by tumor size, cancer-specific mortality rates, and competing causes of death for 40-49 years old

Let $\pi_{i,x}(t)$ be the proportion of cancer patients with tumor size i (i.e., $<$ 1 cm, 1-2 cm, 2-3 cm, 3-5 cm, and 5+ cm.). These proportions can be computed by age, (e.g., ages 40-49 and 50 or older) so that $\pi_i(t) = \pi_{i,40-49}(t) + \pi_{i,50+}(t)$.

$$\sum_{i=1}^{5} \pi_i(t) \quad = \quad \sum_{i=1}^{5} \pi_{i,40-49}(t) + \pi_{i,50+}(t) = 1 \tag{0.3}$$

Then, the change in life expectancy at age x between times t_1 and t_2 can be estimated as:

$$\begin{aligned}
e(x,t_2) - e(x,t_1) &= \sum_{i=1}^{5} \left[\pi_i(t_2)\, e_i(x,t_2) - \pi_i(t_1)\, e_i(x,t_1)\right] \\
&= \sum_{i=1}^{5} \{[\pi_{i,40-49}(t_2) + \pi_{i,50+}(t_2)]\, e_i(x,t_2) - [\pi_{i,40-49}(t_1) + \pi_{i,50+}(t_1)]\, e_i(x,t_1)\} \\
&= \sum_{i=1}^{5} \{\pi_{i,40-49}(t_2)\, e_i(x,t_2) - \pi_{i,40-49}(t_1)\, e_i(x,t_1)\} + \sum_{i=1}^{5} \{\pi_{i,50+}(t_2)\, e_i(x,t_2) - \pi_{i,50+}(t_1)\, e_i(x,t_1)\}
\end{aligned}$$

Each summation in the above equation can be written as (Kitawaga (1955)):

$$\begin{aligned}
&e(x,t_2) - e(x,t_1) = \\
&\sum_{i=1}^{5} [\pi_{i,40-49}(t_2) - \pi_{i,40-49}(t_1)] \left[\frac{e_i(x,t_1) + e_i(x,t_2)}{2}\right] + \sum_{i=1}^{5} [e_i(x,t_2) - e_i(x,t_1)] \left[\frac{\pi_{i,40-49}(t_1) + \pi_{i,40-49}(t_2)}{2}\right] + \\
&\quad \sum_{i=1}^{5} [\pi_{i,50+}(t_2) - \pi_{i,50+}(t_1)] \left[\frac{e_i(x,t_1) + e_i(x,t_2)}{2}\right] + \sum_{i=1}^{5} [e_i(x,t_2) - e_i(x,t_1)] \left[\frac{\pi_{i,50+}(t_1) + \pi_{i,50+}(t_2)}{2}\right] \\
&= \sum_{i=1}^{5} [\pi_{i,40-49}(t_2) - \pi_{i,40-49}(t_1)] \left[\frac{e_i(x,t_1) + e_i(x,t_2)}{2}\right] + \sum_{i=1}^{5} [\pi_{i,50+}(t_2) - \pi_{i,50+}(t_1)] \left[\frac{e_i(x,t_1) + e_i(x,t_2)}{2}\right] + \\
&\quad \sum_{i=1}^{5} [e_i(x,t_2) - e_i(x,t_1)] \left[\frac{\pi_i(t_1) + \pi_i(t_2)}{2}\right]
\end{aligned} \tag{0.4}$$

The first two terms of equation (0.4) correspond to the contribution of changes in the share of tumor size among people aged 40-49 and 50+ to changes in cancer life expectancy between times 1 and 2. We can additionally estimate the contribution of cancer-specific mortality rates to changes in life expectancy by age. The last term of (0.4) can be written as (see equation (0.1)):

$$\begin{aligned}
e_i(40,t_2) - e_i(40,t_1) = &\sum_{j=1}^{k} \sum_{s=40}^{49} [L_{s,i,j}(t_2) - L_{s,i,j}(t_1)] \left[\frac{L_{s,i,-j}(t_2) + L_{s,i,-j}(t_1)}{2n}\right] + \\
&\sum_{j=1}^{k} \sum_{s=50}^{\omega} [L_{s,i,j}(t_2) - L_{s,i,j}(t_1)] \left[\frac{L_{s,i,-j}(t_2) + L_{s,i,-j}(t_1)}{2n}\right]
\end{aligned} \tag{0.5}$$

Assuming constant mortality within age intervals

Let $M_{x,x+n}$ represent the mortality rate between ages x and $x+n$. Then

$$l_{x+n} = e^{-\int_x^{x+n} \mu(s)\, ds} = e^{-n\, M_{x,x+n}} \tag{0.6}$$

We can then estimate the person-years lived between ages x and $x+n$ as

$$_nL_x = l_x \int_x^{x+n} e^{-M_{x,x+n}(s-x)} ds = l_x \left(\frac{-1}{M_{x,x+n}} (e^{-n\, M_{x,x+n}} - 1) \right) \tag{0.7}$$

so, if we have 5-year age groups, then equation (0.7) would look like

$$_5L_x = l_x \left(\frac{-1}{M_{x,x+5}} (e^{-5\, M_{x,x+5}} - 1) \right) \tag{0.8}$$

For the last age group (open-ended, say 100+), we can assume there are no person-years lived beyond a certain time (say no more than 10yrs) to compute $_+L_{100}$ as

$$_+L_{100} = l_{100} \left(\frac{-1}{M_{100+}} (e^{-10\, M_{100+}} - 1) \right) \tag{0.9}$$