

# 1

## 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  $\sum_{i=1}^5 \pi_i = 1$ .

The change in life expectancy at age  $x$  between times  $t_1$  and  $t_2$  by tumor sizes can be decomposed as (Kitigawa 1955):

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

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 tumor-size-specific cancer life expectancy,  $e_i(x, t_2) - e_i(x, t_1)$ , is due to improvements in cancer mortality and competing causes of death (non-cancer mortality). Using the approach developed in Beltrán-Sánchez et al. (2008),

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

where  $i$  corresponds to tumor size,  $j$  is cause-specific mortality among patients diagnosed with tumor size  $i$  (e.g.,  $j = 1$  is cancer,  $j = 2$  is cardiovascular, etc.),  $s$  is age,  $\omega$  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:

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

where  $\text{Diff}_e$  is given by (2) evaluated at  $x = 40$ .

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

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 group (e.g., 40-49 year olds) such that  $\pi_i(t) = \sum_{a \in \mathcal{A}} \pi_{i,a}(t)$  and  $\sum_{i=1}^5 \pi_i = 1$ , where  $\mathcal{A}$  is the set of age groups. Then, the change in life expectancy at age  $x$  between times  $t_1$  and  $t_2$  can be estimated as:

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

Each summation in the above equation can be written as (see equation (1)):

$$\begin{aligned} e(x, t_2) - e(x, t_1) &= \sum_{i=1}^5 [\pi_{i,40-49}(t_2) - \pi_{i,40-49}(t_1)] \left[ \frac{e_i(x, t_1) + e_i(x, t_2)}{2} \right] + \sum_{i=1}^5 [e_i(x, t_2) - e_i(x, t_1)] \left[ \frac{\pi_{i,40-49}(t_1) + \pi_{i,40-49}(t_2)}{2} \right] + \\ &\quad \sum_{i=1}^5 [\pi_{i,50+}(t_2) - \pi_{i,50+}(t_1)] \left[ \frac{e_i(x, t_1) + e_i(x, t_2)}{2} \right] + \sum_{i=1}^5 [e_i(x, t_2) - e_i(x, t_1)] \left[ \frac{\pi_{i,50+}(t_1) + \pi_{i,50+}(t_2)}{2} \right] \\ &= \sum_{i=1}^5 [\pi_{i,40-49}(t_2) - \pi_{i,40-49}(t_1)] \left[ \frac{e_i(x, t_1) + e_i(x, t_2)}{2} \right] + \\ &\quad \sum_{i=1}^5 [\pi_{i,50+}(t_2) - \pi_{i,50+}(t_1)] \left[ \frac{e_i(x, t_1) + e_i(x, t_2)}{2} \right] + \\ &\quad \sum_{i=1}^5 [e_i(x, t_2) - e_i(x, t_1)] \left[ \frac{\pi_i(t_1) + \pi_i(t_2)}{2} \right] \quad (3) \end{aligned}$$

The first two terms of equation (3) 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 tumor-size-specific life expectancy by age. The last term of (3) can be written as (see equation (2)):

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

## Assuming constant mortality within age intervals

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

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

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

$${}_nL_x = l_x \int_x^{x+n} e^{-M_{x,x+n}(s-x)} ds = l_x \left( \frac{-1}{M_{x,x+n}} (e^{-n M_{x,x+n}} - 1) \right). \quad (4)$$

If, for example, age intervals are 5 years wide, equation (4) equals

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

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

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