

This is the decomposition you have done so far:

Let  $\pi_i(t)$  and  $e_i(x, t)$  be the proportion of patients and the life expectancy at age  $x$  time  $t$ , respectively, for cancer patients with stage  $i$  (i.e., localized, regional, distant). That is,  $e_i(x, t)$  would be a stage-specific life expectancy. The overall life expectancy at age  $x$  time  $t$  would be

$$e(x, t) = \sum_{i=0}^3 \pi_i(t) e_i(x, t)$$

where  $\sum_{i=0}^3 \pi_i = 1$ .

Then, the change in life expectancy at age  $x$  between times  $t_1$  and  $t_2$  can be decomposed as (Kitagawa 1955):

$$\begin{aligned} e(x, t_2) - e(x, t_1) &= \sum_{i=0}^3 [\pi_i(t_2) e_i(x, t_2) - \pi_i(t_1) e_i(x, t_1)] \\ &= \sum_{i=0}^3 [\pi_i(t_2) - \pi_i(t_1)] \left[ \frac{e_i(x, t_1) + e_i(x, t_2)}{2} \right] + \sum_{i=0}^3 [e_i(x, t_2) - e_i(x, t_1)] \left[ \frac{\pi_i(t_1) + \pi_i(t_2)}{2} \right] \end{aligned} \quad (0.1)$$

The above equation tells us how much of change in life expectancy at age  $x$  between times  $t_1$  and  $t_2$  is due to: a) changes in cancer stage distribution due to shifts in cancer stage (first term) and b) changes in stage-specific life expectancy (mortality).

Formula (0.1) would produce 6 terms (one for each cancer-stage proportion and for each cancer-stage mortality), which are the pieces you plotted as blue and red bars.

The next step is to further decompose the second term in (0.1). The idea would be to estimate how much of the change in cancer-specific life expectancy,  $e_i(x, t_2) - e_i(x, t_1)$ , is due to cancer and non-cancer mortality improvements. In this case we are looking at cause of death  $j$  among patients diagnosed with cancer  $i$ ; this can be done using traditional decomposition approaches as follows (eqn 2 in my 2008 paper):

$$e_i(x, t_2) - e_i(x, t_1) = \sum_{j=1}^k \sum_{x=0}^{\omega} [L_{x,i,j}(t_2) - L_{x,i,j}(t_1)] \left[ \frac{L_{x,i,-j}(t_2) + L_{x,i,-j}(t_1)}{2n} \right] \quad (0.2)$$

where  $i$  corresponds to cancer,  $j$  is cause-specific mortality among patients diagnosed with cancer  $i$ ,  $x$  is age, and  $L_x$  are person-years lived from the life table.

For simplicity, I think we only want to look at two causes of death among cancer patients: deaths due to cancer  $i$  and all-other causes. That is, we are interested in finding the contribution of improvements in cancer mortality  $i$  among those diagnosed with this type of cancer.

Thus, we don't need additional proportions in equation (0.1) for the additional decomposition, we just need to estimate equation (0.2) and use these results in the second term of equation (0.1).