Supplementary Materials for "Quantifying the Contribution of Earlier Detection and Advancements in Treatment on Gains in Life Expectancy for US Breast Cancer Patients Since 1975"

Samir Soneji[*] Hiram Beltrán-Sánchez[†]

[*]Norris Cotton Cancer Center and Dartmouth Institute for Health Policy & Clinical Practice, Geisel School of Medicine at Dartmouth. Email: samir.soneji@dartmouth.edu

[†]Community Health Sciences, University of California, Los Angeles. Email: beltrans@ucla.edu

A Computation of Incidence-Based Case Fatality Rates

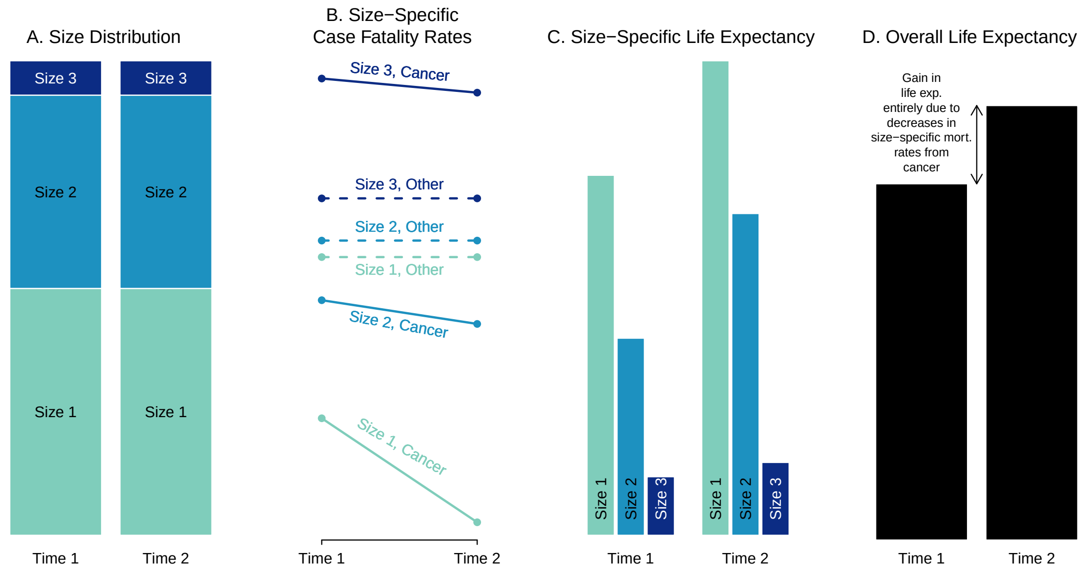
An incidence-based case fatality rate for a specific cohort of newly diagnosed breast cancer patients equals the ratio of [1] the number of deaths occurring for this cohort up to 10 years beyond their diagnosis and [2] the total number of person-years lived by this cohort up to 10 years beyond their diagnosis. For example, 556 women aged 65-69 years were diagnosed with $<$1 cm breast cancer in 2001. Between 2001 and 2011, 22 of these women died of breast cancer and another 107 died of a competing cause of death. This entire cohort lived a total of 5099.5 person-years over the 10-year period. Thus, the incidence-based case fatality rate from breast cancer equaled 22/5099.5 and the incidence-based case fatality rate from competing causes of death equaled 107/5099.5.

B Adjustment for Overdiagnosis

Suppose 10% of the 556 women aged 65-69 years old diagnosed with ¡1cm breast cancer in 2001 were overdiagnosed, the observed case fatality rate from breast cancer (22/11,591) would become 22/[11,591 - 0.10*11,591]. Formally, let $m_{a,t,i}$ represent the observed case fatality rate for age group $a \in \mathcal{A}$ (e.g., 40-44 years, ..., $\geq$100 years), year $t \in \mathcal{T}$ (e.g., 1975, ... 2002), and tumor size $i \in \mathcal{S}$. Let α_i represent the assummed level of overdiagnosis for tumor size i. Then, the case fatality rate adjusted for overdiagnosis, $m^*_{a,t,i}$ equals, $\frac{1}{1-\alpha_i} \times m_{a,t,i}$.

C Schematic Representation of the Methodology

For simplicity, consider three mutually exclusive and exhaustive categories of tumor size: 1, 2, and 3 (e.g., $<$1cm, 1-2cm, and $\geq$2cm). Suppose the distribution of tumor size at cancer diagnosis remains constant between times 1 and 2 (Figure 1, Panel A), tumor size-specific case fatality rates from breast cancer decrease between times 1 and 2 (Figure 1, Panel B), and tumor size-specific case fatality rates from competing causes of death remain constant between times 1 and 2 (Figure 1, Panel B). Tumor size-specific life expectancy increases between times 1 and 2 because tumor size-specific case fatality rates from breast cancer decreased over the time period (Figure 1, Panel C). Overall life expectancy at each time equals the weighted average of tumor size-specific life expectancy, where the weights equal the distribution of tumor sizes at cancer diagnosis at times 1 and 2, respectively. Overall life expectancy grew between times 1 and 2, and this gain was entirely due to decreases in tumor size-specific case fatality rates from breast cancer (Figure 1, Panel D). In actuality, all three aforementioned factors change over time and contribute to the gain in life expectancy. We quantify the individual contribution of each of these three constituent components. We also utilize the same demographic method to further disaggregate these three contributions by age group.



Supplemental Figure 1: Changes in the gain in life expectancy depend on three factors. (A) Changes in the tumor size distribution at cancer diagnosis, (B) tumor size-specific case fatality rates from breast cancer, and (C) tumor size-specific life expectancy.

D Decomposition by Tumor Size and Case Fatality Rates from Breast Cancer and Other Causes of Death

Let $\pi_i(t)$ and $e_i(a,t)$ 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 a and time t, respectively. That is, $e_i(a,t)$ represents tumor-size-specific life expectancy. The overall life expectancy at age a time t is given by

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

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

The change in life expectancy at age a between times t_1 and t_2 by tumor sizes can be decomposed using the methodology of Kitagawa (1955):

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

Equation D.1 quantifies how much of the change in life expectancy at age a 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, $e_i(a,t_2) - e_i(a,t_1)$, is due to improvements in cancer mortality and competing causes of death. Using the approach developed in Beltrán-Sánchez et al. (2008),

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

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.; negative indexes correspond to all but the cause, when $j = 1$ then $-j$ is all but cancer, when $j = 2$ then $-j$ is all but cardiovascular, etc), 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:

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

where $\mathtt{Diff_e}$ is given by (D.2) evaluated at $a = 40$.

E Decomposition by Tumor Size, Case Fatality Rates from Breast Cancer and Other Causes of Death, and Age Group

Let $\pi_{i,a}(t)$ 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 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 (e.g., $a = 1$ represents 40-49 years old, $a = 2$ represents 50-59 years old, $\ldots$, $a = 7$ represents 100+ years old). Then, the change in life expectancy at age a between times t_1 and t_2 can be estimated as:

$$
\begin{aligned}
e(40,t_2) - e(40,t_1) &= \sum_{i=1}^{5} \left[\, \pi_i(t_2)\, e_i(40,t_2) - \pi_i(t_1)\, e_i(40,t_1) \right] \\
&= \sum_{i=1}^{5} \left[\sum_{a=1}^{7} \pi_{i,a}(t_2)\, e_i(40,t_2) - \sum_{a=1}^{7} \pi_{i,a}(t_1)\, e_i(40,t_1) \right] \\
&= \sum_{i=1}^{5} \left[\pi_{i,1}(t_2)\, e_i(40,t_2) - \pi_{i,1}(t_1)\, e_i(40,t_1) \right] + \\
&\quad \sum_{i=1}^{5} \left[\pi_{i,2}(t_2)\, e_i(40,t_2) - \pi_{i,2}(t_1)\, e_i(40,t_1) \right] + \\
&\quad \vdots \\
&\quad \sum_{i=1}^{5} \left[\pi_{i,7}(t_2)\, e_i(40,t_7) - \pi_{i,7}(t_1)\, e_i(40,t_1) \right].
\end{aligned}
$$

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

$$
\begin{aligned}
& e(40,t_2) - e(40,t_1) = \\
& \sum_{i=1}^{5} \left[\pi_{i,1}(t_2) - \pi_{i,1}(t_1)\right] \left[\frac{e_i(40,t_1) + e_i(40,t_2)}{2}\right] + \sum_{i=1}^{5} \left[e_i(40,t_2) - e_i(40,t_1)\right] \left[\frac{\pi_{i,1}(t_1) + \pi_{i,1}(t_2)}{2}\right] + \\
& \sum_{i=1}^{5} \left[\pi_{i,2}(t_2) - \pi_{i,2}(t_1)\right] \left[\frac{e_i(40,t_1) + e_i(40,t_2)}{2}\right] + \sum_{i=1}^{5} \left[e_i(40,t_2) - e_i(40,t_1)\right] \left[\frac{\pi_{i,2}(t_1) + \pi_{i,2}(t_2)}{2}\right] + \\
& \vdots \\
& \sum_{i=1}^{5} \left[\pi_{i,7}(t_2) - \pi_{i,7}(t_1)\right] \left[\frac{e_i(40,t_1) + e_i(40,t_2)}{2}\right] + \sum_{i=1}^{5} \left[e_i(40,t_2) - e_i(40,t_1)\right] \left[\frac{\pi_{i,7}(t_1) + \pi_{i,7}(t_2)}{2}\right] = \\
& \sum_{i=1}^{5} \left[\mathtt{Diff}_{\pi,1}\right] \bar{\mathtt{e}}_{\mathtt{i}} + \sum_{i=1}^{5} \left[\mathtt{Diff}_{\pi,2}\right] \bar{\mathtt{e}}_{\mathtt{i}} + \cdots + \sum_{i=1}^{5} \left[\mathtt{Diff}_{\pi,7}\right] \bar{\mathtt{e}}_{\mathtt{i}} + \sum_{i=1}^{5} \left[e_i(40,t_2) - e_i(40,t_1)\right] \left[\frac{\pi_i(t_1) + \pi_i(t_2)}{2}\right]
\end{aligned}
\tag{E.1}
$$

where $\mathtt{Diff}_{\pi,\mathtt{a}} = \pi_{i,a}(t_2) - \pi_{i,a}(t_1)$ for $a = 1, \ldots, 7$ and $\bar{\mathtt{e}}_{\mathtt{i}} = \frac{e_i(40,t_1) + e_i(40,t_2)}{2}$.

The first seven terms of equation (E.1), those involving $\texttt{Diff}_{\pi,1}$ through $\texttt{Diff}_{\pi,7}$, correspond to the contribution of changes in the share of tumor size by age group to changes in cancer life expectancy between times t_1 and t_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 (E.1) can be written as (see equation (D.2)):

$$
\begin{aligned}
e_i(40,t_2) - e_i(40,t_1) = &\sum_{j=1}^{k}\sum_{s=40}^{49}\left[L_{s,i,j}(t_2) - L_{s,i,j}(t_1)\right]\left[\frac{L_{s,i,-j}(t_2) + L_{s,i,-j}(t_1)}{2n}\right] + \\
&\sum_{j=1}^{k}\sum_{s=50}^{59}\left[L_{s,i,j}(t_2) - L_{s,i,j}(t_1)\right]\left[\frac{L_{s,i,-j}(t_2) + L_{s,i,-j}(t_1)}{2n}\right] + \\
&\vdots \\
&\sum_{j=1}^{k}\sum_{s=100}^{\omega}\left[L_{s,i,j}(t_2) - L_{s,i,j}(t_1)\right]\left[\frac{L_{s,i,-j}(t_2) + L_{s,i,-j}(t_1)}{2n}\right].
\end{aligned}
\tag{E.2}
$$

F Assuming Constant Mortality Within Age Intervals

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

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

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

$$ {}_nL_a = l_a \int_a^{a+n} e^{-M_{a,a+n}(s-a)}ds = l_a\left(\frac{-1}{M_{a,a+n}}(e^{-n\,M_{a,a+n}} - 1)\right). \tag{F.1}$$

If, for example, age intervals are 5 years wide, equation (F.1) equals

$$ {}_5L_a = l_a\left(\frac{-1}{M_{a,a+5}}(e^{-5\,M_{a,a+5}} - 1)\right).$$

For the last age group (e.g., ≥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

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

G Varying Time Intervals Between Diagnosis and Death

Time Interval (Yrs)	Period	Gain in Life Expectancy (Yrs)	Earlier Detection (Yrs)	Advancements in Treatment of Breast Cancer (Yrs)	Advancements in Treatment of Other Diseases (Yrs)
8	1975-2002	11.23	3.15 (28%)	7.07 (63%)	1.03 (9%)
9	1975-2002	10.93	3.09 (28%)	6.76 (62%)	1.09 (10%)
10	1975-2002	10.69	2.99 (28%)	6.57 (61%)	1.15 (11%)
11	1975-2002	10.38	2.78 (27%)	6.27 (60%)	1.35 (13%)
12	1975-2002	10.28	2.65 (26%)	6.05 (59%)	1.59 (15%)

Supplemental Table 1: Gain in Life Expectancy and Contribution from Earlier Detection, Advancements in Breast Cancer Treatment, and Advancements in Treatment of Other Diseases, Varying Incidence-Based Case Fatality Rate Window Length. Note: Yrs=years.

Time Interval	Period	Gain in Life Expectancy	Earlier Detection	Advancements in Treatment of Breast Cancer	Advancements in Treatment of Other Diseases
8	1975-2002	11.23	3.15 (28%)	7.07 (63%)	1.03 (9%)
9	1975-2002	10.93	3.09 (28%)	6.76 (62%)	1.09 (10%)
10	1975-2002	10.69	2.99 (28%)	6.57 (61%)	1.15 (11%)
11	1975-2002	10.38	2.78 (27%)	6.27 (60%)	1.35 (13%)
12	1975-2002	10.28	2.65 (26%)	6.05 (59%)	1.59 (15%)

Supplemental Table 2: Gain in Life Expectancy and Contribution from Earlier Detection, Advancements in Breast Cancer Treatment, and Advancements in Treatment of Other Diseases, Varying Incidence-Based Case Fatality Rate Window Length. Note: Yrs=years.

References

Beltrán-Sánchez, H., Preston, S. H., and Canudas-Romo, V. (2008), “An Integrated Approach to Cause-of-Death Analysis: Cause-Deleted Life Tables and Decompositions of Life Expectancy,” *Demographic Research*, 19, 1323–1350.

Kitagawa, E. (1955), “Components of a Difference Between Two Rates,” *Journal of the American Statistical Association*, 50, 1168–1194.