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. Also, the proportion of women diagnosed with ¡1cm breast cancer in 2001 equaled 4,602 out of 19,029 newly diagnosed breast cancers (24.2%).

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 (22/5099.5) would become 22/[5099.5 - 0.10*5099.5]. Formally, let $\mathcal{A}$ be a set of starting ages for age intervals analyzed (e.g., 40, 45, $\ldots$, $\omega = 100$ years), $\mathcal{T}$ be a set of years (e.g., 1975, $\ldots$2002), and $\mathcal{S}$ be a set of tumor sizes at diagnosis (e.g., $<$ 1cm, 1-2cm, 2-3cm, 3-5cm, and $\geq$ 5cm). Let α_s represent the assummed level of overdiagnosis for tumor size $s \in \mathcal{S}$. Let $m_{a,t,s}$ represent the observed case fatality rate for age group $a \in \mathcal{A}$, year $t \in \mathcal{T}$, and tumor size $s \in \mathcal{S}$. Then, the case fatality rate adjusted for overdiagnosis, $m^*_{a,t,s}$, equals, $\frac{1}{1-\alpha_s} \times m_{a,t,s}$.

In 2001, the number of women diagnosed with breast cancer equaled: 4602 with $<$ 1cm tumors, 7208 with 1-2cm tumors, 3684 with 2-3cm tumors, and 1300 with $\geq$ 5cm tumors. These counts translate to the following distribution: 24%, 38% 19%, 12%, and 7%, respectively. Suppose 10% of $<$ 1cm breast cancers were overdiagnosed (460 of 4602 women). We subtract these 460 women from the count of breast cancers in 2001 and recalculate the distribution: 22% for $<$ 1cm, 39% for 1-2cm, 20% for 2-3cm, 12% for 3-5cm, and 7% for $\geq$ 5cm. Let α_s represent the assummed level of overdiagnosis for tumor size $s \in \mathcal{S}$. Let $n_{t,s}$ represent the observed count of breast cancer cases in year t and for tumor size s. The observed distribution of incident breast cancer cases, $\pi_{t,s}$, equals $\frac{n_{t,s}}{\sum_{s \in \mathcal{S}} n_{t,s}}$. The distribution of incident breast cancer cases adjusted for overdiagnosis equals:

$$\pi^*_{t,s} = \frac{(1-\alpha_s) \times n_{t,s}}{\sum_{s \in \mathcal{S}} (1-\alpha_s) \times n_{t,s}}.$$

C Computation of Tumor Size-Specific Life Expectancy

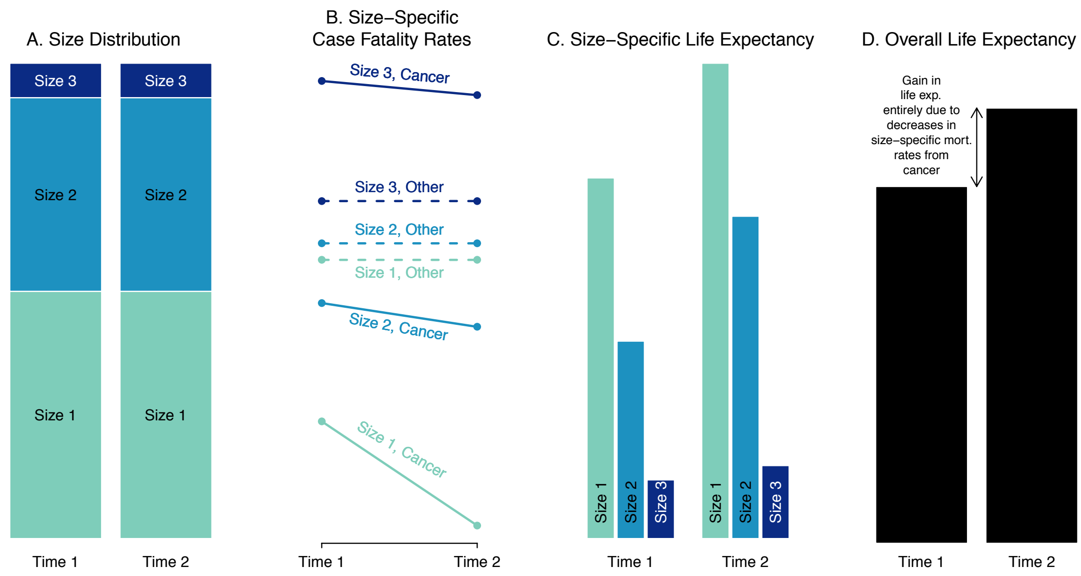
The life expectancy of a breast cancer patient newly diagnosed at age $a^* \in \mathcal{A}$, at time t, and with tumor size s equals:

$$e_s(a^*, t) = \int_{a^*}^{\omega} e^{\left(-\int_{a*}^{a} \mu_s(y,t)\,dy\right)} da = \int_{a^*}^{\omega} e^{\left(-\sum_{a=a*}^{a} n\, m^*_{a,t,s}\right)}\, da, \tag{C.1}$$

where $\mu_s(a,t)$ and $m^*_{a,t,s}$ represent the hazard of mortality and case fatality rate adjusted for overdiagnosis, respectively; n is the width of the age interval; and ω is the starting age of the final and open-ended age interval.

D 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 ≥2cm). Suppose the distribution of tumor size at cancer diagnosis remained constant between times 1 and 2 (Supplemental Figure 1, Panel A), tumor size-specific case fatality rates from breast cancer decreased between times 1 and 2 (Supplemental Figure 1, Panel B), and tumor size-specific case fatality rates from competing causes of death remained constant between times 1 and 2 (Supplemental Figure 1, Panel B). Tumor size-specific life expectancy increased between times 1 and 2 because tumor size-specific case fatality rates from breast cancer decreased over the time period (Supplemental 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 (Supplemental 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 in Section F.



Supplemental Figure 1: The gain in life expectancy depends on three factors: (A) changes in the tumor size distribution at cancer diagnosis, (B) changes in tumor size-specific case fatality rates from breast cancer, and (C) changes in tumor size-specific case fatality rates from competing causes of death.

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

Let $\pi_s(t)$ equal the proportion of breast cancer patients diagnosed with tumor size s in year t. Let $e_s(a,t)$ equal the tumor size-specific life expectancy at age a. The overall life expectancy at age a and time t, $e(a,t)$, equals:

$$e(a,t) = \sum_{s \in \mathcal{S}} \pi_s(t)\, e_s(a,t),$$

where $\sum_{s \in \mathcal{S}} \pi_s = 1$.

The change in life expectancy at age a between times t_1 and t_2 can be decomposed using the methodology of **?**:

$$\begin{aligned} e(a,t_2) - e(a,t_1) &= \sum_{s \in \mathcal{S}} \left[\pi_s(t_2)\, e_s(a,t_2) - \pi_s(t_1)\, e_s(a,t_1)\right] \\ &= \sum_{s \in \mathcal{S}} \left[\pi_s(t_2) - \pi_s(t_1)\right] \left[\frac{e_s(a,t_1) + e_s(a,t_2)}{2}\right] + \\ &\quad \sum_{s \in \mathcal{S}} \left[e_s(a,t_2) - e_s(a,t_1)\right] \left[\frac{\pi_s(t_1) + \pi_s(t_2)}{2}\right]. \end{aligned} \tag{E.1}$$

Equation E.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 of equation E.1 by cause of death. In doing so, we can quantify how much of this change in tumor-size-specific cancer life expectancy, $e_s(a, t_2) - e_s(a, t_1)$, is due to improvements in case fatality rates from breast cancer and improvements in case fatality rates from competing causes of death. Let $\mathcal{C}$ be a set of mutually exclusive and exhaustive causes of death (e.g., breast cancer and all other causes). Let $L_{a,s,c}(t)$ represent the person-years lived in the life table based on the case fatality rate at age a, for tumor size s, from cause $c \in \mathcal{C}$, and at time t. Similarly, let $L_{a,s,-c}(t)$ represent the person-years lived in the life table based on the case fatality rate at age a, for tumor size s, and from causes other than c ($-c$), and at time t. Let a^* be the first starting age of $\mathcal{A}$. Then, following the approach developed by **?**,

$$e_s(a^*, t_2) - e_s(a^*, t_1) = \sum_{c \in \mathcal{C}} \sum_{a=a^*}^{\omega} [L_{a,s,c}(t_2) - L_{a,s,c}(t_1)] \left[\frac{L_{a,s,-c}(t_2) + L_{a,s,-c}(t_1)}{2n} \right], \qquad \text{(E.2)}$$

where n is the width of the age interval and ω is the starting age of the final and open-ended age interval.

We perform the decomposition starting at age 40; the final decomposition equation equals:

$$\begin{aligned} e(40, t_2) - e(40, t_1) &= \sum_{s \in \mathcal{S}} [\, \pi_s(t_2)\, e_s(40, t_2) - \pi_s(t_1)\, e_s(40, t_1)] \\ &= \sum_{s \in \mathcal{S}} [\pi_s(t_2) - \pi_s(t_1)] \left[\frac{e_s(40, t_1) + e_s(40, t_2)}{2} \right] + \sum_{s \in \mathcal{S}} [\mathtt{Diff_e}] \left[\frac{\pi_s(t_1) + \pi_s(t_2)}{2} \right], \end{aligned}$$

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

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

As previously defined $\pi_s(t)$ equals the proportion of cancer patients with tumor size s in year t. This proportion can also be computed by age group such that $\pi_s(t) = \sum_{a \in \mathcal{A}} \pi_{s,a}(t)$ and $\sum_{s \in \mathcal{S}} \pi_s = 1$. Then, the change in life expectancy at age a between times t_1 and t_2 can be estimated as:

$$
\begin{aligned}
e(40,t_2) - e(40,t_1) &= \sum_{s\in\mathcal{S}} \left[\, \pi_s(t_2)\, e_s(40,t_2) - \pi_s(t_1)\, e_s(40,t_1)\right] \\
&= \sum_{s\in\mathcal{S}} \left[\sum_{a\in\mathcal{A}} \pi_{s,a}(t_2)\, e_s(40,t_2) - \sum_{a\in\mathcal{A}} \pi_{s,a}(t_1)\, e_s(40,t_1) \right] \\
&= \sum_{s\in\mathcal{S}} \left[\pi_{s,40}(t_2)\, e_s(40,t_2) - \pi_{s,40}(t_1)\, e_s(40,t_1)\right] + \\
&\quad \sum_{s\in\mathcal{S}} \left[\pi_{s,45}(t_2)\, e_s(40,t_2) - \pi_{s,45}(t_1)\, e_s(40,t_1)\right] + \\
&\quad \vdots \\
&\quad \sum_{s\in\mathcal{S}} \left[\pi_{s,\omega}(t_2)\, e_s(40,t_\omega) - \pi_{s,\omega}(t_1)\, e_s(40,t_1)\right].
\end{aligned}
$$

Each summation in the above equation can be written as follows based on equation (E.1):

$$
\begin{aligned}
&e(40,t_2) - e(40,t_1) = \\
&\sum_{s\in\mathcal{S}} [\pi_{s,40}(t_2) - \pi_{s,40}(t_1)] \left[\frac{e_s(40,t_1) + e_s(40,t_2)}{2}\right] + \sum_{s\in\mathcal{S}} [e_s(40,t_2) - e_s(40,t_1)] \left[\frac{\pi_{s,40}(t_1) + \pi_{s,40}(t_2)}{2}\right] + \\
&\sum_{s\in\mathcal{S}} [\pi_{s,45}(t_2) - \pi_{s,45}(t_1)] \left[\frac{e_s(40,t_1) + e_s(40,t_2)}{2}\right] + \sum_{s\in\mathcal{S}} [e_s(40,t_2) - e_s(40,t_1)] \left[\frac{\pi_{s,45}(t_1) + \pi_{s,45}(t_2)}{2}\right] + \\
&\vdots \\
&\sum_{s\in\mathcal{S}} [\pi_{s,\omega}(t_2) - \pi_{s,\omega}(t_1)] \left[\frac{e_s(40,t_1) + e_s(40,t_2)}{2}\right] + \sum_{s\in\mathcal{S}} [e_s(40,t_2) - e_s(40,t_1)] \left[\frac{\pi_{s,\omega}(t_1) + \pi_{s,\omega}(t_2)}{2}\right] = \\
&\sum_{s\in\mathcal{S}} [\mathtt{Diff}_{\pi,\mathtt{40}}]\, \bar{\mathtt{e}}_{\mathtt{s}} + \sum_{s\in\mathcal{S}} [\mathtt{Diff}_{\pi,\mathtt{45}}]\, \bar{\mathtt{e}}_{\mathtt{s}} + \cdots + \sum_{s\in\mathcal{S}} [\mathtt{Diff}_{\pi,\omega}]\, \bar{\mathtt{e}}_{\mathtt{s}} + \sum_{s\in\mathcal{S}} [e_s(40,t_2) - e_s(40,t_1)] \left[\frac{\pi_s(t_1) + \pi_s(t_2)}{2}\right]
\end{aligned}
\tag{F.1}
$$

where $\mathtt{Diff}_{\pi,\mathtt{a}} = \pi_{s,a}(t_2) - \pi_{s,a}(t_1)$ and $\bar{\mathtt{e}}_{\mathtt{i}} = \frac{e_s(40,t_1)+e_s(40,t_2)}{2}$.

The terms of equation(F.1) that include $\mathtt{Diff}_{\pi,\mathtt{40}} \ldots \mathtt{Diff}_{\pi,\omega}$ correspond to the contribution of changes in the share of tumor size by age group to the change in cancer life expectancy between times t_1 and t_2. We can additionally estimate the contribution of changes in case fatality rates from breast cancer and competing causes of death to changes in tumor-size-specific life expectancy by age. The last term of (F.1) can be written as follows, based on equation (E.2):

$$
e_s(40,t_2) - e_s(40,t_1) = \sum_{c\in\mathcal{C}} \sum_{a=40}^{\omega} [L_{a,s,c}(t_2) - L_{a,s,c}(t_1)] \left[\frac{L_{a,s,-c}(t_2) + L_{a,s,-c}(t_1)}{2n}\right] \tag{F.2}
$$

.

G Assuming Constant Mortality Within Age Intervals

Let $M_{a,a+n}$ represent the mortality rate between ages a and $a+n$ and let $\mu(a)$ represent the hazard of mortality at age a. Then, the number (or proportion) of survivors at age $a+n$ in the life table, l_{a+n}, equals (**?**):

$$l_{a+n} = l_a \, e^{-\int_a^{a+n} \mu(x)\, dx} = l_a \, e^{-n\, M_{a,a+n}}.$$

Then, the number of person-years lived between ages a and $a+n$ equals:

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

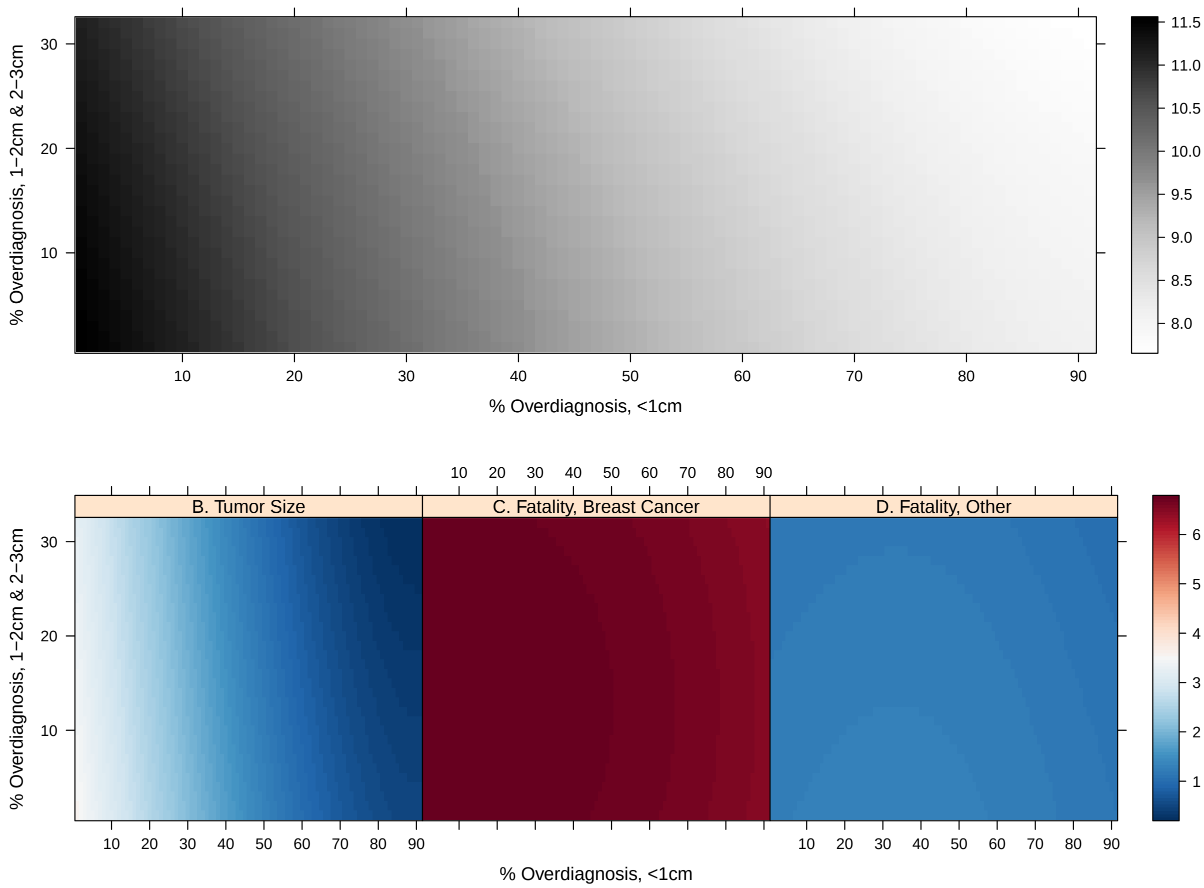
Suppose the age interval is $n = 5$ years wide, then equation (G.1) equals:

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

For the last and open-ended 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 $_\infty L_{100}$ as:

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

H Varying Overdiagnosis Level for < 1cm and $1 - 3$cm Tumors



Supplemental Figure 2: Gain in life expectancy (top panel) and contributions of the temporal shift to smaller sized tumors (bottom left), temporal reductions in case fatality rates from breast cancer (bottom center), and temporal reductions in case fatality rates from competing causes of death (bottom right) varying the overdiagnosis level for < 1cm tumors (0% to 90%) and 1-3cm tumors (0% to 31%). The color scale for the top (bottom) panel indicates the number of years of the gain in life expectancy (contribution to the gain).

I Varying Time Intervals Between Diagnosis and Death

Time Interval	Gain in Life Expectancy	Tumor Size	Reductions in Case Fatality Rates from Breast Cancer	Competing Causes
8	11.23	3.15 (28%)	7.07 (63%)	1.03 (9%)
9	10.93	3.09 (28%)	6.76 (62%)	1.09 (10%)
10	10.69	2.99 (28%)	6.57 (61%)	1.15 (11%)
11	10.38	2.78 (27%)	6.27 (60%)	1.35 (13%)
12	10.28	2.65 (26%)	6.05 (59%)	1.59 (15%)

Supplemental Table 1: Gain in life expectancy and contribution of the temporal shift to smaller sized tumors, temporal reductions in case fatality rates from breast cancer, and temporal reductions in case fatality rates from competing causes of death, 1975-2000, varying time interval between breast cancer diagnosis and death. Note: Yrs=years.