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 $\mathcal{A}$ be a set of age intervals (e.g., 40-44 years, ..., ≥100 years), $\mathcal{T}$ be a set of years (e.g., 1975, ...2002), and $\mathcal{S}$ be a set of tumor sizes at diagnosis(e.g., < 1 cm, 1-2 cm, 2-3 cm, 3-5 cm, and 5+ cm). 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, $\frac{1}{1-\alpha_s} \times m_{a,t,s}$.

C Computation of Tumor size-specific Life Expectancy

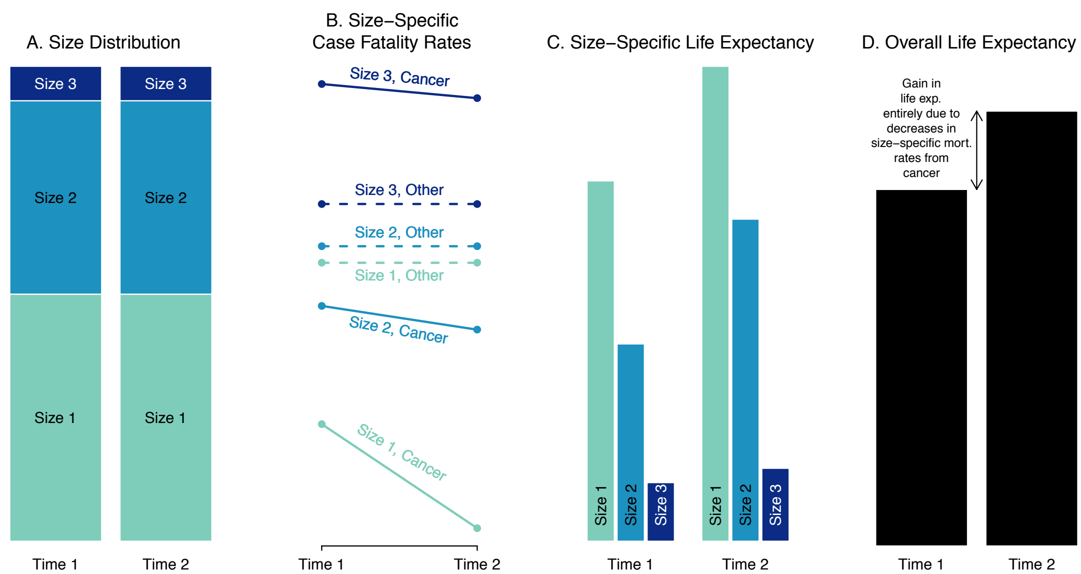
Life expectancy at age a^*, time t, and tumor size s can be estimated as

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

where $\mu_s(a, t)$ and $m^*_{a,t,s}$ represent the hazard of mortality and observed case fatality rate, respectively, for age group $a \in \mathcal{A}$, year $t \in \mathcal{T}$, and tumor size $s \in \mathcal{S}$; 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 remains constant between times 1 and 2 (Figure 2, Panel A), tumor size-specific case fatality rates from breast cancer decrease between times 1 and 2 (Figure 2, Panel B), and tumor size-specific case fatality rates from competing causes of death remain constant between times 1 and 2 (Figure 2, 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 2, 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 2, 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.

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

Let $\pi_s(t)$ and $e_s(a,t)$ be the proportion of patients and the life expectancy for cancer patients with tumor size s, at age a, and at time t. That is, $e_s(a,t)$ represents tumor size-specific life expectancy. The overall life expectancy at age a time t is given by

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

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

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

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

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 of equation D.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 cancer mortality and 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 $c \in \mathcal{C}$, 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 $(-c)$, at time t. Then, following the approach developed by Beltrán-Sánchez et al. (2008),

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

where n is the width of the age interval.

Samir: I don't think we can simplify the second summation by stating that $a \in \mathcal{A}$. The reason is that we are decomposition life exp at age a, so this summation must start at this age and end at ω. I think we need to use another letter to index the summation; something like

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

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

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

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

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

Let $\pi_s(t)$ be 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:

$$\begin{aligned} e(40,t_2) - e(40,t_1) &= \sum_{s\in\mathcal{S}} \left[\, \pi_s(t_2)\, e_s(40,t_2) - \pi_s(t_1)\, e_s(40,t_1)\right] \\ &= \sum_{s\in\mathcal{S}} \left[\sum_{a\in\mathcal{A}} \pi_{s,a}(t_2)\, e_s(40,t_2) - \sum_{a\in\mathcal{A}} \pi_{s,a}(t_1)\, e_s(40,t_1)\right] \\ &= \sum_{s\in\mathcal{S}} \left[\pi_{i,1}(t_2)\, e_s(40,t_2) - \pi_{i,1}(t_1)\, e_s(40,t_1)\right] + \\ &\quad \sum_{s\in\mathcal{S}} \left[\pi_{i,2}(t_2)\, e_s(40,t_2) - \pi_{i,2}(t_1)\, e_s(40,t_1)\right] + \\ &\quad \vdots \\ &\quad \sum_{s\in\mathcal{S}} \left[\pi_{i,|\mathcal{A}|}(t_2)\, e_s(40,t_{|\mathcal{A}|}) - \pi_{i,|\mathcal{A}|}(t_1)\, e_s(40,t_1)\right]. \end{aligned}$$

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

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

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 first $|\mathcal{A}|$ terms of equation (E.1) (the terms based on $\mathtt{Diff}_{\pi,1}$ through $\mathtt{Diff}_{\pi,|\mathcal{A}|}$) 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 follows, based on equation (D.2):

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

Samir: Same comment as above, I don't think we can simplify the second summation by stating that $a\in\mathcal{A}$. This summation must start at age 40 and end at ω. I suggest we write it as

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

G Assuming Constant Mortality Within Age Intervals

Let $M_{a,a+n}$ represent the mortality rate between ages a and $a+n$. Then, we can estimate the life table survivors at age $a+n$ as (Preston et al. 2001)

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

where $\mu(x)$ represents the risk of death at age x.

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

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

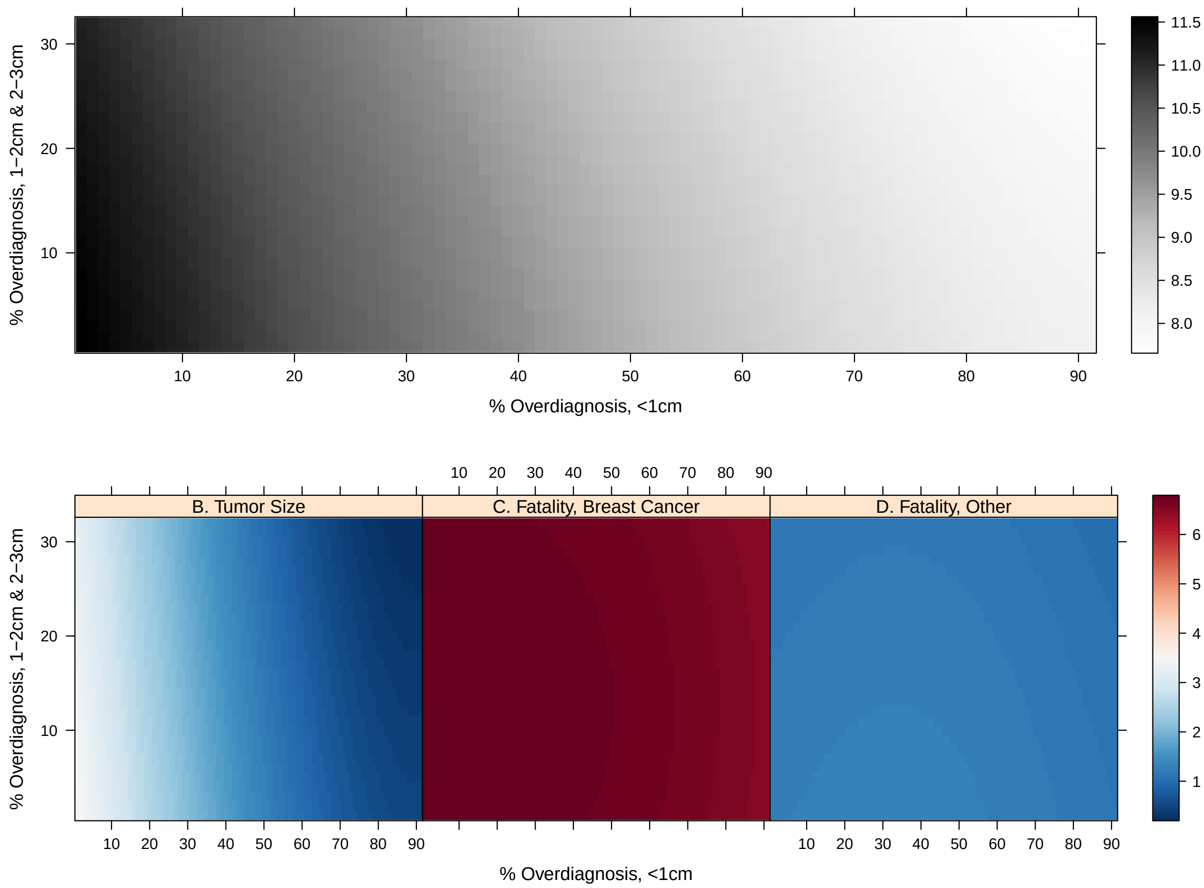
If, for example, age intervals are 5 years wide, equation (F.1) equals

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

For the last age group (e.g., $\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

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

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



Supplemental Figure 2: Gain in life expectancy (top panel) and contributions of earlier detection (bottom left), advancements in breast cancer treatment (bottom center), and advancements in treatment of other diseases (bottom right) varying overdiagnosis level for < 1cm tumors (0% to 90%) and 1-3cm tumors (0% to 31%).

I Varying Time Intervals Between Diagnosis and Death

Time		Gain in	Earlier	Advancements in Treatment of	
nterval	Period	Life Expectancy	Detection	Breast Cancer	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 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.

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.

Preston, S. H., Heuveline, P., and Guillot, M. (2000), *Demography: Measuring and Modeling Population Processes*, Oxford, UK: Blackwell.