THE LARGE UNOBSERVED CASE RATE OF CANCER USING BACKWARD INDUCTION METHOD REVEALS SCREENING OPPORTUNITY IN EARLIER STAGES

EPH232

Andrew K. ElHabr[1], Christopher Tyson[2], Xiting Cao[2], Selin Merdan[1], Sana Raoof[3], A. Burak Ozbay[2], Paul Limburg[2], Tomasz M. Beer[2], Ashish Deshmukh[4], Jagpreet Chhatwal[1,5,6]

[1]Value Analytics Labs, Boston, MA, USA, [2]Exact Sciences Corporation, Madison, WI, USA, [3]Memorial Sloan Kettering Cancer Center, New York, NY, USA, [4]Medical University of South Carolina, Charleston, SC, USA, [5]Massachusetts General Hospital, Boston, MA, USA, [6]Dana Farber / Harvard Cancer Center, Boston, MA

BACKGROUND

- Cancer is the second leading cause of death in the US and the world [1].
- In 2021, there were 1.9 million new cases of cancer in the US [2].
- However, population-level cancer registries report diagnosed incident cases only, and underlying unobserved cancer burden is unknown.
- An estimate of the unobserved case rate of cancer is important to inform policies and interventions (including new screening technologies) that can increase early detection.

OBJECTIVES

- To develop a method to estimate the unobserved case rate of cancer using mathematical modeling and available data from cancer registries.
- To apply this method to infer the hidden number of cases of two cancer types: pancreatic and colorectal cancer.

METHODS

Rationale: any advanced stage cancer at some point was in an early stage and remained undiagnosed. By backtracking advanced stage cancers to early stages, we can estimate the (unobserved) case rate for each cancer stage.

Table 1: *Mathematical notation for methodology.*

Symbol	Definition	Examples
s	Sex	Female, Male
a	Age group	50-54, 55-59, …, 80-84, 85+
c	Cancer type	Colorectal, Pancreatic
t	Year	2010, 2011, …, 2015
$\mu_{s,a}$	Annual mortality rate for sex s and age group a	(NVSS [4])
$m_{s,a}$	Quarterly survival probability for sex s and age group a	(Step 2)
$d(c,j \to k)$	Cumulative dwell time for cancer type c from stage j to stage k; $j,k \in \{1,2,3,4\}$ where $j < k$	(see Table 2 for dwell times between consecutive stages)
$f(m_{s,a}, d(c,j \to k))$	"Background mortality inflation factor" for sex s and age group a using cumulative dwell time for cancer type c from stage j to stage k	(Step 3)
$O_{c,j,s,a,t}$	Stage j observed case rate of cancer type c for sex s and age group a in year t	(SEER-17 [3])
$U_{c,j,s,a,t}$	Stage j unobserved case rate of cancer type c for sex s and age group a in year t	(Step 3)
$I_{c,j,t}$	Stage j total (observed + unobserved) case rate of cancer type c in year t	(Step 5)
$w_{s,a}$	The proportion of people of sex s and in age group a among the age 50+ US population	(ACS [5])

KEY FINDINGS

1. We used a novel approach to estimate undiagnosed cancer case rate in the US. We used the rationale that any advanced stage cancer at some point was in an early stage and remained undiagnosed.
2. Our approach suggests a large undiagnosed case rate of colorectal and pancreatic cancers, especially in early stages.
3. Using this novel approach, we can estimate the undiagnosed case rate for all major cancers.
4. Understanding the unknown cancer case rate is important when assessing the economic and population-level impact of screening and diagnostic technologies for early cancer detection.

Specifically, we use data collected from the National Cancer Institute's Surveillance, Epidemiology, and End Results database (SEER-17) [3], CDC's National Vital Statistics Systems (NVSS) [4], and the US Census Bureau's American Community Survey (ACS) [5]. The methodology is as follows:

1. Interpolate age- and year-case rates to reflect rates by quarter of a year.
2. For each sex s and age group a pairing, generate quarterly survival probability ($m_{s,a}$) using annual mortality rate ($\mu_{s,a}$) from NVSS.

$$m_{s,a} = \left(1 - \mu_{s,a}\right)^{1/4}$$

3. Apply backwards induction equations using cumulative dwell times (the average time it takes for cancer type c to progress from stage j to stage k), which were synthesized in an ensemble approach using input from published literature and empirical estimates (Table 2), to estimate stagewise unobserved case rates ($U_{c,j,s,a,t}$, in which stage is indexed as j) of each cancer type c for each sex s and age group a in year t from known stagewise observed case rates ($O_{c,j,s,a,t}$). Cumulative dwell time of cancer type c to progress from stage j to stage k is denoted as $d(c,j \to k)$. Additionally, each observed case rate must be "inflated" to account for any competing background mortality that occurred over time using a factor we label the "background mortality inflation factor".

$$f\left(m_{s,a}, d(c, j \to k)\right) = m_{s,\lfloor a+0.25 \rfloor} \times m_{s,\lfloor a+0.5 \rfloor} \times \cdots \times m_{s,\lfloor a+d(c,j \to k)-0.25 \rfloor} \times m_{s,\lfloor a+d(c,j \to k) \rfloor}$$

$$U_{c,3,s,a,t} = \frac{O_{c,4,s,a+d(c,3\to4),t+d(c,3\to4)}}{f\left(m_{s,a}, d(c, 3 \to 4)\right)}$$

$$U_{c,2,s,a,t} = \frac{O_{c,3,s,a+d(c,2\to3),t+d(c,2\to3)}}{f\left(m_{s,a}, d(c, 2 \to 3)\right)} + \frac{O_{c,4,s,a+d(c,2\to4),t+d(c,2\to4)}}{f\left(m_{s,a}, d(c, 2 \to 4)\right)}$$

$$U_{c,1,s,a,t} = \frac{O_{c,2,s,a+d(c,1\to2),t+d(c,1\to2)}}{f\left(m_{s,a}, d(c, 1 \to 2)\right)} + \frac{O_{c,3,s,a+d(c,1\to3),t+d(c,1\to3)}}{f\left(m_{s,a}, d(c, 1 \to 3)\right)} + \frac{O_{c,4,s,a+d(c,1\to4),t+d(c,1\to4)}}{f\left(m_{s,a}, d(c, 1 \to 4)\right)}$$

4. Compute the stage j unobserved case rates ($U_{c,j,t}$) of cancer type c in year t among the age 50+ US population using a weighted sum of sex-age grouping proportions, with weights $w_{s,a}$ obtained from the ACS in which $\sum_{s,a} w_{s,a} = 1$.

$$U_{c,j,t} = \sum_{s,a} w_{s,a}\, U_{c,j,s,a,t}$$

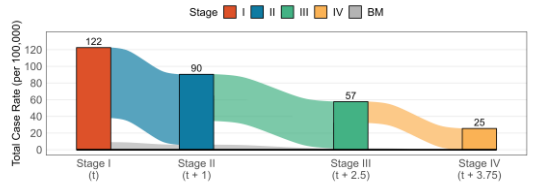
5. Compute the stage j total case rates ($I_{c,j,t}$) of each cancer type c in year t (Figure 1).

$$I_{c,j,t} = O_{c,j,t} + U_{c,j,t}$$

Table 2: *Dwell times by cancer between consecutive stages in terms of years. For example, the stage I dwell time for colorectal cancer implies that it takes 1 year on average for the cancer to progress from stage I to stage II. Similarly, the stage III dwell time for colorectal cancer implies that it takes 1.25 years on average for the cancer to progress from stage III to stage IV. The dwell times between consecutive stages are used to estimate cumulative dwell times between stages. For example, the cumulative dwell times for colorectal cancer between stage I and stage III is 1 + 1.5 = 2.5 years.*

Cancer Type	Stage I	Stage II	Stage III	Stage IV
Colorectal	1	1.5	1.25	0.75
Pancreatic	1	1	0.75	0.5

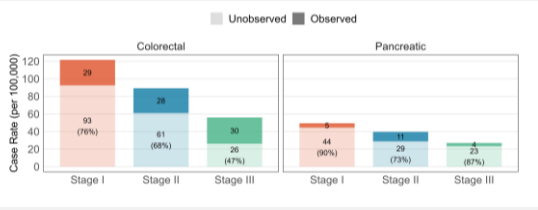
Figure 1: *Backwards induction computation for the stage I total case rate of colorectal cancer in 2015. (BM: background mortality).*



RESULTS

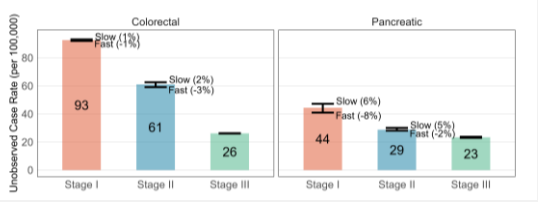
- The following results are for the age 50+ US population in 2015.
- The proportions of stage I, II, and III cancer cases that remain undetected are (Figure 2):
  - **pancreatic**: 90%, 73%, 87%;
  - **colorectal**: 76%, 68%, 47%.

Figure 2: *Estimated unobserved (and observed) case rates for the age 50+ US population in 2015 with the proportion of total cases that remain undetected.*



- Thus, for each identified stage I **pancreatic** and **colorectal** cancer case our methodology analyses yields there to be **9** and **3** additional cases that remained undetected.
- The sensitivity of the estimated unobserved case rates to dwell time scenarios in which all dwell times are decreased/increased by a factor of 30% to reflect "fast"- and "slow"-developing scenarios is minimal (Figure 3).

Figure 3: *Sensitivity of the estimated unobserved case rates for the age 50+ US population in 2015 in "fast"- and "slow"-developing scenarios.*



REFERENCES

[1] Mohsen Naghavi, et al. "Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016". The Lancet 390. 10100(2017): 1151–1210.

[2] "Cancer Facts & Figures 2021". American Cancer Society, https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf. Accessed 18 April 2023.

[3] Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Plus Data, 17 Registries, Nov 2021 Sub (2000-2019) - Linked To County Attributes - Total U.S., 1969-2020 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2022, based on the November 2021 submission.

[4] Arias, Elizabeth, and Jiaquan Xu. "United States Life Tables, 2018." National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System vol. 69,12 (2020): 1-45.

[5] "Age and Sex." 2015 American Community Survey 1-Year Estimates, United States Census Bureau, 2023, https://data.census.gov/table?tid=ACSST5Y2017.S0101. Accessed 10 January 2023.

CONTACT Andrew K. ElHabr, PhD / *andrewelhabr@valueanalyticslabs.com*