National Cancer Registrars Association

ISSN: 1945-6123

Journal of Registry Management

Winter 2022 | Volume 49 | Number 4

Published by the National Cancer Registrars Association
Founded in 1975 as *The Abstract*

The Right Job for the Ideal Candidate

JOB BANK IS NCRA'S RESOURCE CONNECTING EMPLOYERS TO CANCER REGISTRARS.

OPPORTUNITY AWAITS

visit Job Bank today at

www.ncra-usa.org/jobs

Journal of Registry Management

is published quarterly by the
National Cancer Registrars Association
1330 Braddock Place, #520
Alexandria, VA 22314
(703) 299-6640
(703) 299-6620 FAX

Address change of address and subscription correspondence to:
National Cancer Registrars Association
1330 Braddock Place, #520
Alexandria, VA 22314

Address all editorial correspondence to:
Neisha Johnson
Email: JRMEditor@ncra-usa.org
Telephone: (703) 299-6640 ext. 319

Letters to the Editor
Letters to the Editor must be signed and include address and telephone number; none will be published anonymously. Letters subject to editing.

Editorial Policy
Opinions expressed in articles appearing in *Journal of Registry Management* are those of the authors and not necessarily those of the National Cancer Registrars Association, or the editor of *Journal of Registry Management*.

Subscription to *Journal of Registry Management*
Subscription is a benefit of membership in the National Cancer Registrars Association. Individual subscriptions may be purchased for $40/year U.S. dollars. Single copies may be purchased for $12 U.S., $15 International, prepaid only.

Advertising Sales
Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the *Journal* does not indicate or imply endorsement by the *Journal* or NCRA. Advertising sales should be directed to Marketing Coordinator, NCRA, 1330 Braddock Place, #520, Alexandria, VA 22314, (703) 299-6640, (703) 299-6620 FAX.

National Cancer Registrars Association

Journal of Registry Management

Winter 2022 • Volume 49 • Number 4

Contents

Continued on page 107

Indexing

The *Journal of Registry Management* is indexed in the National Library of Medicine's MEDLINE database. Citations from the articles indexed, the indexing terms (key words), and the English abstract printed in *JRM* are included and searchable using PubMed.

For your convenience, the *Journal of Registry Management* is indexed in the 4th issue of each year and on the Web (under "Resources" at http://www.ncra-usa.org/jrm). The 4th issue indexes all articles for that particular year. The Web index is a cumulative index of all *JRM* articles ever published.

Journal of Registry Management

is published quarterly by the
National Cancer Registrars Association
1330 Braddock Place, #520
Alexandria, VA 22314
(703) 299-6640
(703) 299-6620 FAX

Address change of address and subscription correspondence to:
National Cancer Registrars Association
1330 Braddock Place, #520
Alexandria, VA 22314

Address all editorial correspondence to:
Neisha Johnson
Email: JRMEditor@ncra-usa.org
Telephone: (703) 299-6640 ext. 319

Letters to the Editor
Letters to the Editor must be signed and include address and telephone number; none will be published anonymously. Letters subject to editing.

Editorial Policy
Opinions expressed in articles appearing in *Journal of Registry Management* are those of the authors and not necessarily those of the National Cancer Registrars Association, or the editor of *Journal of Registry Management*.

Subscription to *Journal of Registry Management*
Subscription is a benefit of membership in the National Cancer Registrars Association. Individual subscriptions may be purchased for $40/year U.S. dollars. Single copies may be purchased for $12 U.S., $15 International, prepaid only.

Advertising Sales
Advertising is accepted with the understanding that products and services are professionally related and meet the ethical standards of public health practice. Acceptance by the *Journal* does not indicate or imply endorsement by the *Journal* or NCRA. Advertising sales should be directed to Marketing Coordinator, NCRA, 1330 Braddock Place, #520, Alexandria, VA 22314, (703) 299-6640, (703) 299-6620 FAX.

National Cancer Registrars Association

Journal of Registry Management

Winter 2022 • Volume 49 • Number 4

Contents, continued

Letter from the Editor

Dear Colleagues,

I am pleased to present the third annual North American Association of Central Cancer Registries) NAACCR special edition of the *Journal of Registry Management.* This annual special edition provides opportunities for NAACCR members to feature their cancer surveillance research or related work. I encourage NAACCR members to plan to submit articles again next fall. In addition to original articles, I encourage registries to submit short reports or an editorial viewpoint on an issue impacting their registry for consideration next year.

This year's special issue includes 9 original articles, 1 short report, and the 3 winning posters from the NAACCR 2022 Virtual Forum. The papers presented underwent a peer-review process overseen by NAACCR Research and Data Use Steering Committee members. Poster authors participated in the Call for Abstracts for the NAACCR Summer Forum held in June 2022. After initial peer review, the authors of submissions accepted as a poster had the opportunity to include their posters in an online judging session. Ultimately, 3 posters were selected for awards.

Published papers include 2 papers that focus different aspects of the impact of COVID-19: one on early COVID-19 hospitalization among patients with cancer history (Zhang, et al), and one on using claims data to get an early look at the decrease in incident cancer cases (Zhang, et al), which is the subject of the Winter 2023 Continuing Education Quiz. Other articles include an assessment of linking with LexisNexis to support a residential mobility study (Tatalovich, et al); an evaluation of Do Not Contact codes (Lawson-Michod, et al); integrating screening data into routine cancer surveillance (Hernandez, et al); the introduction of a new variable to support analysis based on clinically relevant brain and central nervous system tumors (Ostrom, et al); a discussion of informatics initiatives impacting registries (Rollison, et al); factors associated with loss to follow-up (Qiao, et al); and determining fitness for use in research of the SEER cause specific cause of death (Morawski, et al). We also have a short report on prostate cancer out of Montana (Zimmerman).

The second-place winning poster also focuses on the impact of COVID-19, specifically racial and ethnic disparities of COVID-19 among women with precancerous cervical lesions (Hsieh, et al). The first-place poster presents on comparisons among participants and nonparticipants in patient contact studies (Kuliszewski, et al), and the third-place poster describes a childhood cancer study in New Hampshire (Ricci, et al).

Once again, it has been a privilege to collaborate with *JRM* on this third publication of NAACCR focused articles. Please note, the findings and conclusions in this report are those of the authors and do not necessarily represent the views of the NAACCR or the JRM.

Be well,

Recinda Sherman, PhD, CTR

Guest Editor, *Journal of Registry Management*

Original Article

Assessment of Interstate Residential Mobility of SEER Patients: SEER and LexisNexis Residential Address Linkage

Zaria Tatalovich, PhD[a]; David G. Stinchcomb, MS, MA[b]; Angela Mariotto, PhD[a]; Diane Ng, MPH[b]; Jennifer L. Stevens, BS[c]; Linda M. Coyle, BS[c]; Lynne Penberthy, MD, MPH[a]

Abstract: The National Cancer Institute (NCI) Surveillance, Epidemiology, and End Results (SEER) program is continuously exploring opportunities to augment its already extensive collection of data, enhance the quality of reported cancer information, and contribute to more comprehensive analyses of cancer burden. This manuscript describes a recent linkage of the LexisNexis longitudinal residential history data with 11 SEER registries and provides estimates of the inter-state mobility of SEER cancer patients. To identify mobility from one state to another, we used state postal abbreviations to generate state-level residential histories. From this, we determined how often cancer patients moved from state-to-state. The results in this paper provide information on the linkage with LexisNexis data and useful information on state-to-state residential mobility patterns of a large portion of US cancer patients for the most recent 1-, 2-, 3-, 4-, and 5-year periods. We show that mobility patterns vary by geographic area, race/ethnicity and age, and cancer patients tend to move less than the general population.

Key words: *data linkage; exposure estimates; residential history; social determinants; Surveillance, Epidemiology, and End Results (SEER) program*

Introduction

The National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results (SEER) program is a rich source of cancer related information including diagnostic data, patient demographics, tumor characteristics, initial treatment at the time of diagnosis, and outcomes.[1] The SEER Program is continuously exploring opportunities to augment its already extensive collection of data, enhance the quality of reported cancer information, and contribute to more comprehensive analysis of cancer burden. The emerging sources of cancer-related data coupled with novel technologies for data extraction and linkage present an opportunity for cancer registries to integrate larger-scale longitudinal data pre-and post-diagnosis into the existing cancer surveillance data infrastructure.

While cancer registries collect the patient's residential address at the time of diagnosis, historical and updated address histories are not generally available. Having residential history pre- and post- cancer diagnosis would facilitate data linkages with multiple sources of longitudinal data, enhance the quality of data linkage in the absence of patient identifier information, and provide research opportunities to investigate the association of exposures to neighborhood social and environmental conditions with risks of developing cancer over the life course[2-7] as well as the impact of a cancer diagnosis on cancer survivorship issues.[2,8-10] For example, incorporating residential history records into cancer research can enhance our understanding of the impacts of neighborhood sociodemographic and physical conditions, poverty and social deprivation, accessibility to healthcare resources, quality and availability of cancer care, tobacco and alcohol consumption, food environments, and contaminants in water, soil, and air at various places of residence on cancer risk and outcomes Once diagnosed with cancer, patients may move for a variety of reasons: to be closer to their families, for better access to treatment, or for other survivorship considerations,[9] or as a result of losing a job due to poor health or disability. Until recently, individual residential history data have been difficult and expensive to obtain. Studies requiring residential history records for cancer patients often relied on patient's self-reported addresses, introducing recall bias with no means of assessing this error, or incomplete addresses derived from electronic medical records, introducing collection bias.[11] Increasingly, commercial resources of residential history data such as LexisNexis[12] offer easier access to, and more complete, individual address information, which presents an opportunity for cancer control research community to reconstruct residential histories of cancer patients.

In 2016, NCI sponsored a pilot study to assess the accuracy and completeness of residential history data provided by three vendors including LexisNexis, compared to self-reported address from 66 volunteer participants at NCI and NIEHS who represented varying age and migratory history. Of the three vendors, LexisNexis was identified as a source of the most complete, accurate, and available residential history data dating back to the 1980s.[13] Other studies, limited to a single registry, conducted assessment of LexisNexis residential history data[6,11] and concluded that LexisNexis address records can be used for reconstructing residential histories in cancer surveillance and epidemiological research.

[a] Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, Maryland. [b] Westat, Inc, Rockville, Maryland. [c] Information Management Services (IMS), Inc, Calverton, Maryland.

Address correspondence to Zaria Tatalovich, PhD, 9609 Medical Center Dr, Rockville, MD 20850. Email: tatalovichzp@mail.nih.gov.

This manuscript describes a recent enhanced linkage of the LexisNexis longitudinal residential history data with 11 SEER registries and provides estimates of the inter-state mobility of SEER cancer patients based on this linkage. Because most data received by cancer registries are within the state, knowing how often cancer patients move out of the state of diagnosis can inform the percent of patients that may not be linking to state data. To our knowledge, no study has investigated the inter-state mobility patterns of a large population based database of cancer patients.

Methods

Linkage

LexisNexis maintains a commercially available database containing information from a variety of data sources on more than 276 million US individuals.[12] Based on the prior linkage with LexisNexis,[13] 11 SEER registries (10 state registries and one metropolitan-area registry (Seattle)) who had already established confidentiality agreements with LexisNexis were included in this study. We included cancer patients who were at least 21 years old and had been diagnosed between 2009 and 2015 because the residential history data for younger ages and earlier diagnoses years were not as complete. Death certificate only cases were excluded since only limited address information is available for these cases. The cohort included approximately 3,247,000 cancer patients. For each cancer patient in the cohort, the following data items were sent to LexisNexis to conduct the linkage: first name, middle name, last name, suffix, Social Security Number (SSN), address at diagnosis (street, city, state and zip code), date of birth and phone number. The linkage was conducted in 2019. The percentage of cases in the SEER data with a complete SSN was approximately 96%.

Developing Residential Histories and Conducting State-to-State Mobility Analysis

Data returned by LexisNexis included any address associated with an individual and a range of dates when that address was used. The data often contained multiple records for the same residence with minor differences, multiple unique residence records for overlapping time periods, or a gap in residence records during the time period. To construct each patient's residential history, i.e. a single address at any particular time point, the data needs to be reconciled and adjusted for overlaps and gaps in addresses. To identify mobility from one state to another, we used state postal abbreviations which are rarely misspelled and can be easily reconciled to generate state-level residential histories. From this, we determined how often cancer patients moved from state-to-state. For their final state of residence, we determined the number of years in this state and noted patients that moved to a different state within 1 year, 2 years, 3 years, 4 years, and 5 years. From this we calculated the state-level move rates as the percent of patients who have moved to a different state within the most recent number of years. Note that this time period varies for each patient depending on the end date of the most recent address returned by LexisNexis. These time periods are looking backwards in time from the most recent residence reported by LexisNexis and, thus, include residence periods both before and after the date of diagnosis. For this study, we looked only at the LexisNexis address data, so we were not able to differentiate between pre- and post-diagnosis locations.

For the United States, data on the residential mobility of the general population is available from the Census Bureau[14] and these data have been analyzed for older adults.[15] We used 5-year data from the Census Bureau's American Community Survey for 2015-2019 to calculate state-level move rates for the general population stratified by geographic area, sex, race/ethnicity and age group. Since the cancer patients are generally older than the general population and previous studies have shown that older adults move frequently, we used age group profiles of the cancer population to create weighted state-level move rates. These rates provide estimates of the state-level move rates for a subset of the general population with matching age profiles.

Results

As shown in Table 1, LexisNexis was able to link and return address information on 3,117,258 (98.5%) of the patients sent for linkage. We received up to the maximum of 20 address records for each patient, with an average of 7.7 records per patient. The percentage linked by registry was highest for Connecticut, Georgia, Kentucky, Louisiana, Seattle, and Utah (over 99%) and lowest for New York (97.7%). Linkage rates were very similar by sex but were lower for the non-Hispanic Asian and Pacific Islander API (94.3%) and Hispanic (95.7%) patients. By age at first diagnosis, linkage rates were highest for those diagnosed between 50 and 64 years (98.9%) and were lowest for patients diagnosed at the youngest (97.5%) and oldest (97.1%) age groups. By diagnosis year, linkage rates were very similar.

The percentage of cancer patients who moved to a different state within the most recent 1 year, 2 years, 3 years, 4 years, and 5 years is shown in Table 2 for 11 SEER registry areas. About 1 percent or less of cancer patients have moved to a different state within the most recent 1 year; whereas between 2.5 and 4.7 percent have moved within the last 5 years. Cancer patients in New York have the most state-to-state moves and patients in Louisiana have the least. Among cancer patients in these registries, females move from state-to-state a bit more often than males. By race/ethnic groups, non-Hispanic API patients move from state-to-state the most frequently with non-Hispanic White patients moving the least often. As expected, younger patients move from state-to-state more often than older patients.

For comparison, Table 2 includes state-level move rates for the general population. The unweighted state-level move rates of the general population are generally higher that the state-level move rates for the cancer population. The weighted move rates which estimate the state-level move rates for a subset of general population with matching age profiles are also generally higher than those for cancer patients. By registry area, the exceptions are the states of

Table 1. SEER Residential History Data Linkage Results by Registry, Demographic Characteristics, and Diagnosis Year

	No. patients submitted	*No. linked and returned with address information (%)*
Total	3,226,404	3,177,258 (98.5)
Registry		
California	1,093,698	1,072,072 (98.0)
Connecticut	149,405	148,344 (99.3)
Georgia	332,737	330,892 (99.4)
Iowa	123,331	122,072 (99.0)
Idaho	54,252	53,694 (99.0)
Kentucky	186,414	185,233 (99.4)
Louisiana	172,361	171,319 (99.4)
New Mexico	64,062	63,245 (98.7)
New York	792,594	774,250 (97.7)
Seattle	184,198	183,138 (99.4)
Utah	73,352	72,999 (99.5)
Sex		
Male	1,559,511	1,537,859 (98.6)
Female	1,666,450	1,638,969 (98.4)
Other/unknown	443	430 (97.1)
Race/ethnicity		
NH White	2,309,662	2,292,590 (99.3)
NH Black	335,079	329,553 (98.4)
NH AI/AN	14,152	14,024 (99.1)
NH API	184,377	173,879 (94.3)
Hispanic	346,317	331,587 (95.7)
Unknown	36,817	35,625 (96.8)
Age at first diagnosis (y)		
20-24	18,386	17,919 (97.5)
25–29	34,396	33,719 (98.0)
30–34	50,703	49,818 (98.3)
35–39	70,902	69,765 (98.4)
40–44	118,189	116,448 (98.5)
45–49	187,984	185,493 (98.7)
50–54	289,516	286,192 (98.9)
55–59	370,108	365,919 (98.9)
60–64	437,590	432,652 (98.9)
65–69	457,842	452,263 (98.8)
70–74	384,928	379,158 (98.5)
75–79	318,363	312,503 (98.2)
80–84	249,814	244,727 (98.0)
≥85	237,674	230,677 (97.1)
Unknown	9	5 (55.6)
Diagnosis year		
2009	466,879	459,564 (98.4)
2010	458,222	451,186 (98.5)
2011	461,623	454,724 (98.5)
2012	456,053	449,496 (98.6)
2013	457,223	450,333 (98.5)
2014	460,335	453,408 (98.5)
2015	466,069	458,547 (98.4)

AI/AN, American Indian and Alaska Native; API, Asian/Pacific Islander; NH, non-Hispanic.

California and New York where cancer patients have higher state-to-state move rates than their counterparts in the general public. By race/ethnicity, the exceptions are non-Hispanic API and Hispanic cancer patients.

Discussion

This paper demonstrates the feasibility of obtaining residential histories for almost all adult cancer patients diagnosed in recent years in SEER. In addition, this is the first large-scale assessment of the state-to-state mobility patterns of US cancer patients covering 30% of the US population and can provide some initial insights into how often cancer patients move between states for different geographic areas. Knowledge of state-to-state move patterns for cancer patients plays an important role for understanding the need to include out of state data in data linkages. For example, requests for supplemental prescription drug data for a given state registry can include data from neighboring states with significant move rates.

There is some geographic variation in the state-to-state move rates with New York rates being the highest and Louisiana rates being the lowest. There is also some variation by race/ethnicity with non-Hispanic API rates the highest and non-Hispanic White rates the lowest. Older cancer patients move less frequently than younger patients. This is consistent with previous studies that indicate older adults move less frequently.[15] Comparison of the state-to-state move rates of cancer patients with that of the general public show generally lower rates for cancer patients. However, state-to-state move rates were in general very low and under 5%, indicating that less than 5% of cancer patients will be missed in state specific data linkages.

This study has limitations. We focused on state level moves as a first step to identify the need to acquire and link with out-of-state data. We only included 11 registries

Table 2. State-Level Move Rates for Cancer Patients by Registry and by Demographic Characteristics for the Most Recent 1-Year to 5-Year Periods with Comparative 1-Year Move Rates for the General Population

Percent of cancer patients who moved to a different state within the most recent N years[a]						*One-year state move rates for general population*[b]	
	1 y	**2 y**	**3 y**	**4 y**	**5 y**	**Unweighted**	**Weighted**[3]
Registry							
California	0.95	1.6	2.4	3.2	3.9	1.30	0.78
Connecticut	0.78	1.5	2.3	3.3	4.2	2.31	1.21
Georgia	0.77	1.4	2.0	2.8	3.5	2.75	1.78
Idaho	0.77	1.5	2.3	3.2	4.2	4.33	3.14
Iowa	0.51	1.0	1.5	2.1	2.7	2.50	1.16
Kentucky	0.57	1.0	1.6	2.1	2.8	2.46	1.28
Louisiana	0.50	0.9	1.4	2.0	2.5	1.70	1.00
New Mexico	0.75	1.5	2.4	3.3	4.3	2.92	2.08
New York	1.06	1.9	2.7	3.7	4.7	1.34	0.62
Seattle	0.80	1.6	2.4	3.2	4.1	3.49	1.88
Utah	0.72	1.5	2.3	3.2	4.1	3.26	2.29
Sex							
Male	0.80	1.4	2.1	2.8	3.6	1.94	1.80
Female	0.93	1.7	2.5	3.4	4.2	1.81	1.95
Race/ethnicity							
NH White	0.71	1.3	2.0	2.7	3.4	2.18	2.22
NH Black[d]	0.85	1.5	2.3	3.1	4.0	1.97	1.87
NH API[d]	2.00	3.1	4.2	5.5	7.0	1.87	1.52
NH AI/AN[d]	0.78	1.4	2.1	2.8	3.5	1.75	1.71
Hispanic	1.30	2.1	3.0	4.2	5.3	1.07	0.79
Age at diagnosis (y)							
20–24	2.45	5.3	8.3	11.7	15.2	4.24	4.26
25–29	2.16	4.6	7.1	9.8	12.6	3.93	3.96
30–34	1.62	3.5	5.4	7.6	9.7	2.78	2.80
35–39	1.28	2.6	4.0	5.6	7.2	1.98	1.99
40–44	1.15	2.1	3.3	4.5	5.8	1.46	1.45
45–49	0.98	1.9	2.9	3.9	4.8	1.18	1.17
50–54	0.91	1.7	2.6	3.5	4.4	1.07	1.07
55–59	0.85	1.6	2.4	3.3	4.1	0.98	0.99
60–64	0.81	1.5	2.2	3.1	3.8	0.98	0.99
65–69	0.74	1.3	2.0	2.7	3.4	0.92	0.94
70–74	0.73	1.2	1.8	2.4	3.0	0.82	0.83
≥75	0.78	1.2	1.6	2.1	2.7	0.81	0.79

1. Source: state-level residential history of cancer patients included in the SEER-LN linkage ages 21 and older diagnosed between 2009 and 2015.
2. Source: Census American Community Survey moves from a different state within the last year, 5 year results 2015-2019.
3. Census results are weighted by the age-group profiles of the cancer patients in each of the registry areas.
4. Bridged race/ethnicity categories for non-Hispanic (NH) Black, NH API, and NH AIAN are not available in Census tables. Because of this., move rates for NH Black cancer patients are compared with the single-race Black population of any Hispanic origin; NH API with single-race API of any Hispanic origin, and NH AIAN with single-race AIAN of any Hispanic origin.

representing 30% of the US population. A recent study showed that LexisNexis address information near the time of death may not be accurate.[16]

The results in this paper provide information on the linkage with LexisNexis data and useful information on state-to-state residential mobility patterns of a large portion of US cancer patients for the most recent 1-year, 2-years, 3-years, 4-years, and 5-years. Mobility patterns vary by geographic area, race/ ethnicity and age. Finally, cancer patients tend to move less than the general population.

Work is currently being done to develop an algorithm to construct detailed residential histories that identify unique addresses for a patient with a single address at any particular time point. Once the complete residential history data is created and validated, it will be a unique and valuable resource for extending our understanding of the residential mobility of cancer patients throughout the cancer control continuum as well as providing research opportunities to investigate the association of exposures on outcomes.

Acknowledgement

We would like to acknowledge members of the SEER/ LexisNexis residential address linkage working group: Mary Charlton (Iowa Cancer Registry), Iona Cheng (Greater Bay Area Cancer Registry), Rosemary Cress (Cancer Registry of Greater California), Dennis Deapen (Los Angeles County Cancer Surveillance Program), Will Howe (Information Management Services, Inc.), Tina Lefante (Louisiana Tumor Registry), and Bożena Morawski (Cancer Data Registry of Idaho).

References

1. National Cancer Institute's Surveillance, Epidemiology, and End Results Program website. Accessed December 7, 2022. **https://seer.cancer.gov/**
2. Namin S, Zhou Y, Neuner J, Beyer K. The role of residential history in cancer research: a scoping review. *Soc Sci Med*. 2021;270:113657. doi:10.1016/j.socscimed.2020.113657
3. Tatalovich Z, Wilson JP, Mack T, Yan Y, Cockburn M. The objective assessment of lifetime cumulative ultraviolet exposure for determining melanoma risk. *J Photochem Photobiol B*. 2006;85(3):198-204. doi:10.1016/j.jphotobiol.2006.08.002
4. Paulu C, Aschengrau A, Ozonoff D. Exploring associations between residential location and breast cancer incidence in a case-control study. *Environ Health Perspect*. 2002;110(5):471-478. doi:10.1289/ehp.02110471
5. Jacquez GM, Kaufmann A, Meliker J, Goovaerts P, AvRuskin G, Nriagu J. Global, local and focused geographic clustering for case-control data with residential histories. *Environ Health*. 2005;4(1):4. doi:10.1186/1476-069x-4-4
6. Wheeler DC, Wang A. Assessment of residential history generation using a public-record database. *Int J Environ Res Public Health*. 2015;12(9):11670-11682.
7. Wheeler DC, Waller LA, Cozen W, Ward MH. Spatial-temporal analysis of non-Hodgkin lymphoma risk using multiple residential locations. *Spat Spatiotemporal Epidemiol*. 2012;3(2):163-171. doi:10.1016/j.sste.2012.04.009
8. Wiese D, Stroup AM, Maiti A, et al. Residential mobility and geospatial disparities in colon cancer survival. *Cancer Epidemiol Biomarkers Prev*. 2020;29(11):2119-2125.
9. Liu B, Lee FF, Boscoe F. Residential mobility among adult cancer survivors in the United States. *BMC Pub Health*. 2020;20(1):1-11.
10. Gomez SL, Shariff-Marco S, DeRouen M, et al. The impact of neighborhood social and built environment factors across the cancer continuum: current research, methodological considerations, and future directions. *Cancer*. 2015;121(14):2314-2330. doi:10.1002/cncr.29345
11. Jacquez GM, Slotnick MJ, Meliker JR, AvRuskin G, Copeland G, Nriagu J. Accuracy of commercially available residential histories for epidemiologic studies. *Am J Epidemiol*. 2011;173(2):236-243.
12. LexID. LexisNexis Risk Solutions website. Accessed April 15, 2022. **https://risk.lexisnexis.com/our-technology/lexid**
13. Stinchcomb DG, Roeser A. *NCI/SEER Residential History Project: Technical Report*. Westat, Inc; 2016. **https://www.westat.com/sites/default/files/NCISAS/NCI_Res_Hist_Proj_Tech_Rpt_v2sec.pdf**
14. Frost R. *Are Americans Stuck in Place? Declining Residential Mobility in the US*. Joint Center for Housing Studies of Harvard University; 2020. **https://www.jchs.harvard.edu/sites/default/files/harvard_jchs_are_americans_stuck_in_place_frost_2020.pdf**
15. Choi JH, Goodman L, Zhu J, Walsh J. *Senior Housing and Mobility: Recent Trends and Implications for The Housing Market*. Urban Institute; 2019. **https://www.urban.org/sites/default/files/publication/100953/senior_housing_and_mobility.pdf**
16. Woolpert KM, Ward KC, England CV, Lash TL. Validation of LexisNexis Accurint in the Georgia Cancer Registry's Cancer Recurrence and Information Surveillance Program. *Epidemiology*. 2021;32(3):434-438.

Original Article

Early COVID-19 Hospitalizations Among New York State Residents with a History of Invasive Cancer

Xiuling Zhang, PhD[a]; Margaret Gates Kuliszewski, ScD[a,b]; Amy R. Kahn, MS[a]; Maria J. Schymura, PhD[a,b]

Abstract: Background: Individuals with a history of cancer may be more susceptible to severe COVID-19 due to immunosuppression, comorbidities, or ongoing treatment. We linked inpatient claims data on COVID-19 hospitalizations to cancer diagnoses from the New York State Cancer Registry (NYSCR) to examine associations between prior cancer diagnoses and hospitalizations for COVID-19, and factors associated with death at discharge after COVID-19 hospitalization. Methods: New York State (NYS) residents diagnosed with invasive cancer before July 1, 2021, who were alive on January 1, 2020, were identified from NYSCR data. We obtained claims data for discharge year 2020 and the first half of 2021 from NYS's Statewide Planning and Research Cooperative System (SPARCS), and we linked inpatient records with COVID-19 as the primary diagnosis to cancer data from the NYSCR using deterministic matching methods. We calculated descriptive statistics and conducted multivariable-adjusted logistic regression analyses to examine associations of cancer case characteristics with COVID-19 hospitalization and with vital status at discharge among patients with a history of cancer. All analyses were conducted in SAS 9.4. Results: Our analysis included 1,257,377 individuals with a history of cancer, 10,210 of whom had a subsequent primary COVID-19 hospitalization. Individuals with a history of cancer were 16% more likely to be hospitalized with COVID-19, compared to the general population of NYS, after adjusting for age and sex (95% CI, 14%–19%). Factors independently associated with COVID-19 hospitalization among cancer patients included older age, male sex, non-Hispanic Black race or Hispanic ethnicity, diagnosis with late-stage cancer or with multiple tumors, more recent cancer diagnosis, and New York City (NYC) residency at the time of cancer diagnosis. Factors independently associated with death at discharge among individuals with COVID-19 hospitalization and a prior cancer diagnosis included older age, male sex, non-Hispanic Black or non-Hispanic Asian/Pacific Islander race or Hispanic ethnicity, residence in NYC at the time of COVID-19 hospitalization, and an active cancer diagnosis claim code at the time of COVID-19 hospitalization. Conclusion: This claims-based study identified higher risks of COVID-19 hospitalization and death at discharge among individuals with a history of cancer, and particularly those in certain demographic and diagnostic groups.

Key words: *Cancer, claims, COVID-19, hospitalization, New York State*

Introduction

Cancer is a major public health issue worldwide. In the United States, there were an estimated 18.1 million individuals with any history of cancer (excluding basal cell or squamous cell skin cancers and in situ cancers other than urinary bladder) as of January 1, 2022, representing approximately 5.4% of the total US population.[1-2] Individuals with a history of cancer may be more susceptible to severe COVID-19 due to immunosuppression, comorbidities, or ongoing treatment. Cancer is one of several underlying medical conditions that is considered to be conclusively associated with higher risk for severe COVID-19, based on a review of the literature by the Centers for Disease Control and Prevention (CDC).[3-4]

However, previous studies investigating the relationship between past cancer diagnoses and severe COVID-19 outcomes, including hospitalization and death, have had mixed results. Several studies observed that patients with a cancer history had a significantly higher risk for severe COVID-19 outcomes,[5-12] while others did not[13-14] or only found that patients with a recent cancer diagnosis or those who had received recent cancer treatment were at a higher risk of severe outcomes of COVID-19.[15-16]

The New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) is a comprehensive all-payer database that captures patient-level information for all encounters from hospital inpatient and outpatient visits, making it a valuable resource for data on COVID-19 hospitalizations in NYS.[17] We conducted a comprehensive analysis of early COVID-19 hospitalizations among NYS residents with a history of cancer to examine in detail associations between cancer history and severe COVID-19 outcomes. We linked claims data on COVID-19 hospitalizations from SPARCS to cancer diagnoses from the New York State Cancer Registry (NYSCR) to assess which patient demographics, tumor characteristics, and cancer types were associated with an elevated risk for COVID-19

[a]New York State Cancer Registry, New York State Department of Health, Albany, New York. [b]Department of Epidemiology and Biostatistics, University at Albany School of Public Health, Rensselaer, New York.

Address correspondence to Margaret Gates Kuliszewski, ScD, New York State Cancer Registry, New York State Department of Health, 150 Broadway, Suite 361, Albany, NY 12204. Telephone: (518) 474-2255. Email: maggie.gateskuliszewski@health.ny.gov.

This work was supported in part by the Centers for Disease Control and Prevention's National Program of Cancer Registries through cooperative agreement 6NU58DP006309 awarded to the New York State Department of Health and by Contract 75N91018D00005 (Task Order 75N91018F00001) from the National Cancer Institute, National Institutes of Health.

hospitalization. We then examined associations between prior cancer diagnosis and death at discharge among all COVID-19 hospitalizations obtained from SPARCS, as well as the demographics, tumor characteristics, and cancer types that were associated with an increased risk of death at discharge.

Methods

Data Sources and Case Selection

We retrieved data for 1,262,264 patients with a history of cancer (including 142,114 with more than 1 tumor) from the NYSCR Surveillance, Epidemiology, and End Results Data Management System (SEER*DMS) database for NYS residents who were diagnosed with invasive cancer before July 1, 2021, and who were alive on January 1, 2020. We then obtained claims data from SPARCS, which includes patient-level data on diagnoses from hospital inpatient and outpatient (ambulatory, emergency department, and outpatient services) visits. We retained data for inpatient claims only for NYS residents with discharge dates in 2020 and the first half of 2021 to allow for identification of hospitalizations related to COVID-19. We retrieved a total of 2,977,403 inpatient claims records meeting these inclusion criteria.

Since there is no direct identifier in the SPARCS data, we defined an individual using the date of birth (DOB), sex, and a unique personal identifier (UPI) variable, which is a combination of the first 2 and last 2 letters of the last name, the first 2 letters of the first name, and the last 4 digits of the Social Security number (SSN). When more than 1 claims record had the same UPI, DOB, and sex, we considered them to be the same patient. If the SSN component of UPI was missing, the claims were considered to be from the same patient if they had the same first 6 characters of the UPI and the same DOB, sex, and either patient zip code or both treating facility and medical record number (MRN). From the 2,977,403 inpatient claims records obtained from SPARCS, we identified a total of 2,041,781 unique patients.

Linkage Between Cancer Data and SPARCS Claims Records

We linked SPARCS inpatient records and cancer data using deterministic matching methods by comparing UPI, DOB, sex, reporting facility identifier (PFI), MRN, and patient zip code at diagnosis. The linkage process included 9 sequential steps, followed by manual review to resolve duplicate matches. The 9 steps included linkage of records with: (1) same UPI, DOB, and sex; (2) same UPI, DOB, PFI, and MRN; (3) same UPI, sex, PFI, and MRN; (4) same UPI, DOB, and zip code; (5) same UPI and DOB; (6) same UPI, sex, and either same birth year or same birth month and day; (7) same UPI without SSN, plus same DOB, sex, PFI and MRN; (8) same UPI without SSN, plus same PFI and MRN; and (9) same UPI without SSN, plus same DOB, sex, and zip code.

Of the 1,262,264 cancers retrieved from SEER*DMS, we excluded 4,715 with unknown or implausible age (defined as >110 years), where age was calculated as the difference between the patient's date of birth and either the earliest admission date from SPARCS or, for unlinked cases, the date of the midpoint of the study period (September 1, 2020). We additionally excluded 172 cases with nonmale/nonfemale sex, due to the small number of these cases. After these exclusions, a total of 1,257,377 cancers were included in the study. Of these, 251,304 (20.0%) matched with 1 or more SPARCS inpatient claims records, among which we identified 30 duplicate matches where 2 cancer cases matched to the same inpatient claims record(s). After manual review, we removed 1 case from each duplicate. We also removed 3,878 ineligible matches because their date of cancer diagnosis was later than the date of admission from inpatient claims records, indicating that they had not been diagnosed with cancer prior to the time of their hospitalization. The remaining 247,398 (19.7%) were considered to be good matches, while 1,009,979 cancer cases (80.3%) did not have a documented hospitalization during the time frame of interest.

Identification of COVID-19 Hospitalizations

We identified COVID-19 hospitalizations using *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)* primary diagnosis code U07.1 for discharge year 2020 and the first half of 2021 and code B97.29 for discharge dates from January 1 to March 31, 2020.[18] When a COVID-19 diagnosis code was the principal diagnosis on 1 or more linked inpatient claims records for a patient, it was counted as 1 COVID-19 hospitalization, and only the record with the earliest admission date was included in the analysis. Among the 2,041,781 patients with a hospitalization during the time frame of interest, we identified COVID-19 as a principal diagnosis code for 77,338 cases. Four of these cases had unknown sex and were excluded, leaving a total of 77,334 COVID-19 hospitalizations for analysis, including 10,210 COVID-19 hospitalizations among linked cancer cases.

Statistical Analysis

We calculated descriptive statistics for cancer patients with (vs without) COVID-19 hospitalization for the following demographic and tumor characteristics: age, sex, race/ethnicity, region of residency (New York City [NYC] or the rest of NYS, categorized based on address at cancer diagnosis from SEER*DMS), stage, history of multiple tumors, and time since cancer diagnosis. We used χ^2 tests to examine statistically significant differences in the distribution of each covariate by COVID-19 hospitalization status and student's t test to examine differences in mean age by COVID-19 hospitalization.

We conducted multivariable-adjusted logistic regression analyses to calculate odds ratios (OR) and 95% CIs for associations between the above-mentioned variables and COVID-19 hospitalization among patients with a history of cancer. All variables examined were significantly associated with COVID-19 hospitalization and were adjusted for in the final model. For patients with a history of 2 or more invasive tumors, we included only the most recent invasive tumor in the analysis, based on the assumption that the more recently diagnosed tumor would have had a greater impact on the patient's recent health.

Next, we examined COVID-19 hospitalization by cancer type to determine if patients with a history of certain cancers were more likely to be hospitalized with COVID-19. We categorized cancer diagnoses into 24 cancer types: oral cavity and pharynx; esophagus; stomach; colorectal; liver and intrahepatic bile duct; pancreas; larynx; lung and bronchus; melanoma; breast; cervix uteri; corpus uterus and not otherwise specified (NOS); ovary; prostate; testis; urinary bladder (including in situ); kidney and renal pelvis; brain and other nervous system; thyroid; Hodgkin lymphoma; non-Hodgkin lymphomas; myeloma; leukemias; and other malignancies of hematopoietic or lymphopoietic origin. We included in situ urinary bladder cancers in the analysis based on the SEER rules for determining multiple primary cancers and for calculating incidence rates, which specify that in situ bladder cancers are counted along with invasive cancers when reporting bladder cancer incidence (and for no other type of cancer).

We calculated the ratio of observed to expected (O/E) COVID-19 hospitalizations among cancer cases overall and by cancer type. We estimated the expected counts of COVID-19 hospitalizations using age- and sex-specific rates of COVID-19 hospitalization among NYS residents, which were calculated by dividing age- and sex-specific counts of hospitalizations in NYS from SPARCS by the corresponding age- and sex-specific population counts for NYS from the 2019 American Community Survey 1-year population estimates. We then calculated the expected number of COVID-19 hospitalizations by cancer type by applying these age- and sex-specific proportions of COVID-19 hospitalization for all cancers combined to the observed number of cancers by age and sex for each individual cancer type. We used 18 age groups (0–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, and ≥85 years) and 2 categories of sex (male and female) for these calculations. We calculated 95% CIs for the ratios of observed to expected using the Byar's approximation of the exact Poisson distribution.[19]

Next, we restricted our analysis to all COVID-19 hospitalizations (n = 77,334) retrieved from SPARCS, and we examined vital status at discharge among patients with versus without a history of cancer. In addition, among the patients with a history of cancer, we examined differences in vital status at discharge by the presence or absence of a cancer diagnosis claim code in the SPARCS claim record, where patients with a cancer diagnosis claim code were considered active cancer cases and those without a cancer diagnosis claim code were considered inactive cancer cases. We calculated the crude and age- and sex-adjusted proportions of death for each of these groups, as well as the 95% CIs, using the SAS STDRATE procedure. The age- and sex-adjusted proportions of death by cancer status were calculated using the age- and sex-specific proportions obtained from all 77,334 COVID-19 hospitalization patients as a reference, and the age- and sex-adjusted proportions of death by active status of cancer diagnosis were calculated using the age- and sex-specific proportions obtained from the 10,210 COVID-19 hospitalized patients with a history of cancer as a reference. We used 11 age groups (0–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, and ≥85 years) to adjust for age in these age- and sex-adjusted calculations.

For the 10,210 patients with COVID-19 hospitalization and a history of cancer, we calculated descriptive statistics by vital status at discharge for the following demographic and tumor characteristics: age, sex, race/ethnicity, region of residency at COVID-19 hospitalization, stage, history of multiple tumors, time since cancer diagnosis, and active or inactive cancer diagnosis based on the presence or absence of a cancer claim code at the time of COVID-19 hospitalization. We used χ^2 tests to examine statistically significant differences in the distribution of each covariate by vital status at discharge from hospitalization for COVID-19 and student's *t* test to examine differences in mean age by vital status at discharge. We conducted multivariable-adjusted logistic regression analyses to calculate ORs and 95% CIs for associations between each variable of interest and vital status at discharge. All variables of interest were included as covariates in the final multivariable-adjusted model.

Finally, we used methods similar to those described above to calculate the ratios and 95% CIs of observed to expected counts of deaths at discharge by cancer type among individuals with COVID-19 hospitalization and a history of cancer (n = 10,210). We calculated the expected numbers of deaths at discharge by using age- and sex-specific proportions of death for all 10,210 patients with a history of cancer, based on the data retrieved from SPARCS. We then calculated the expected number of deaths for each cancer type by applying these age- and sex-specific proportions of death at discharge to the observed number of COVID-19 hospitalizations for each cancer type. We calculated 95% CIs for the ratios of observed to expected deaths at discharge based on the Byar's approximation, as described above. All analyses were performed using SAS 9.4.

Results

The overall ratio of observed versus expected COVID-19 hospitalizations among NYS residents with a history of cancer was 1.16 and the 95% CI was 1.14 to 1.19 (results not shown), indicating that individuals with a prior diagnosis of cancer were 16% more likely to be hospitalized with COVID-19 compared to the general population of NYS, after adjusting for age and sex.

Table 1 presents the distribution of demographic and tumor characteristics of interest for individuals with a history of cancer by COVID-19 hospitalization status, as well as multivariable-adjusted ORs and 95% CIs for each variable of interest in relation to COVID-19 hospitalization status. Among individuals with a history of cancer, those with versus without COVID-19 hospitalization tended to be older (mean age of 74.0 vs 68.2 years) and were more likely to be male (55.1% vs 45.9%), non-Hispanic Black (18.9% vs 12.2%), Hispanic (14.3% vs 10.1%), reside in NYC at the time of their cancer diagnosis (45.6% vs 37.7%), and to have been diagnosed with distant-stage or systemic cancer (19.4% vs 11.3%), multiple tumors (16.6% vs 11.2%), and within the past 0 to 2 years (18.0% vs 14.0%) or 3 to 5 years (22.4% vs 19.3%) (all $P < .0001$).

Table 1. Distribution of Demographic and Tumor Characteristics of Interest and Multivariable-Adjusted Odds Ratios (ORs) and 95% CIs for Each Variable of Interest and COVID-19 Hospitalization Status Among 1,257,377 New York State Residents with a History of Cancer

Variable	*Cancer cases with COVID-19 hospitalization (n = 10,210), n (%)*	*Cancer cases without COVID-19 hospitalization (n = 1,247,167), n (%)*	*P value**	*OR (95% CI)***	*OR (95% CI)****
Age in years, mean (SD)	74.0 (12.9)	68.2 (15.2)	<.0001		
Age group (y)			<.0001		
0–19	30 (0.3)	8,340 (0.7)		0.71 (0.49–1.04)	0.71 (0.48–1.03)
20–29	38 (0.4)	14,798 (1.2)		0.63 (0.45–0.88)	0.62 (0.44–0.88)
30–39	71 (0.7)	35,372 (2.8)		0.52 (0.40–0.68)	0.52 (0.40–0.68)
40–49	267 (2.6)	70,967 (5.7)		Ref	Ref
50–59	841 (8.2)	174,431 (14.0)		1.28 (1.12–1.47)	1.27 (1.11–1.46)
60–69	2,041 (20.0)	314,699 (25.2)		1.73 (1.52–1.96)	1.72 (1.51–1.95)
70–79	3,249 (31.8)	348,433 (27.9)		2.59 (2.28–2.93)	2.58 (2.28–2.93)
≥80	3,673 (36.0)	280,127 (22.5)		4.00 (3.53–4.54)	4.01 (3.54–4.55)
Sex			<.0001		
Male	5,627 (55.1)	572,339 (45.9)		1.35 (1.29–1.40)	1.34 (1.29–1.40)
Female	4,583 (44.9)	674,828 (54.1)		Ref	Ref
Race/ethnicity			<.0001		
Non-Hispanic White	6,276 (61.5)	882,587 (70.8)		Ref	Ref
Non-Hispanic Black	1,930 (18.9)	152,041 (12.2)		1.75 (1.66–1.85)	1.91 (1.81–2.01)
Non-Hispanic Asian/Pacific Islander	454 (4.4)	66,127 (5.3)		0.99 (0.89–1.09)	1.09 (0.99–1.20)
Hispanic	1,457 (14.3)	126,327 (10.1)		1.66 (1.56–1.77)	1.83 (1.73–1.94)
Non-Hispanic other/unknown	93 (0.9)	20,085 (1.6)		0.75 (0.61–0.93)	0.80 (0.65–0.98)
Region of residency at cancer diagnosis			<.0001		
New York City	4,655 (45.6)	469,750 (37.7)		1.21 (1.16–1.27)	NA
Rest of New York State	5,555 (54.4)	777,417 (62.3)		Ref	Ref
Stage of Cancer			<.0001		
Local	4,926 (48.2)	626,011 (50.2)		Ref	Ref
Regional	1,628 (15.9)	205,539 (16.5)		1.09 (1.03–1.15)	1.09 (1.03–1.15)
Distant	1,983 (19.4)	141,262 (11.3)		1.69 (1.61–1.79)	1.67 (1.61–1.79)
Unknown stage	1,673 (16.4)	274,355 (22.0)		0.79 (0.75–0.84)	0.80 (0.75–0.84)
Multiple Tumors			<.0001		
No	8,512 (83.4)	1,106,958 (88.8)		Ref	Ref
Yes	1,698 (16.6)	140,209 (11.2)		1.23 (1.16–1.30)	1.23 (1.16–1.29)
Time since cancer diagnosis			<.0001		
0–2 years (2019–2021)	1,840 (18.0)	174,494 (14.0)		1.53 (1.44–1.62)	1.51 (1.42–1.61)
3–5 years (2016–2019)	2,285 (22.4)	240,208 (19.3)		1.42 (1.34–1.50)	1.41 (1.33–1.49)
6–10 years (2012–2016)	2,404 (23.5)	284,264 (22.8)		1.25 (1.19–1.32)	1.25 (1.18–1 .32)
>10 years (1976–2011)	3,681 (36.1)	548,201 (44.0)		Ref	Ref

*P-values from χ^2 tests for categorical variables and student's t test for mean age.
**Analyses mutually adjusted for all variables in the table.
***Analyses mutually adjusted for all variables in the table with the exception of region of residence at cancer diagnosis.

Table 2. Ratio and 95% CI of Observed to Expected Number of COVID-19 Hospitalizations Among Individuals with a History of Cancer by Cancer Type

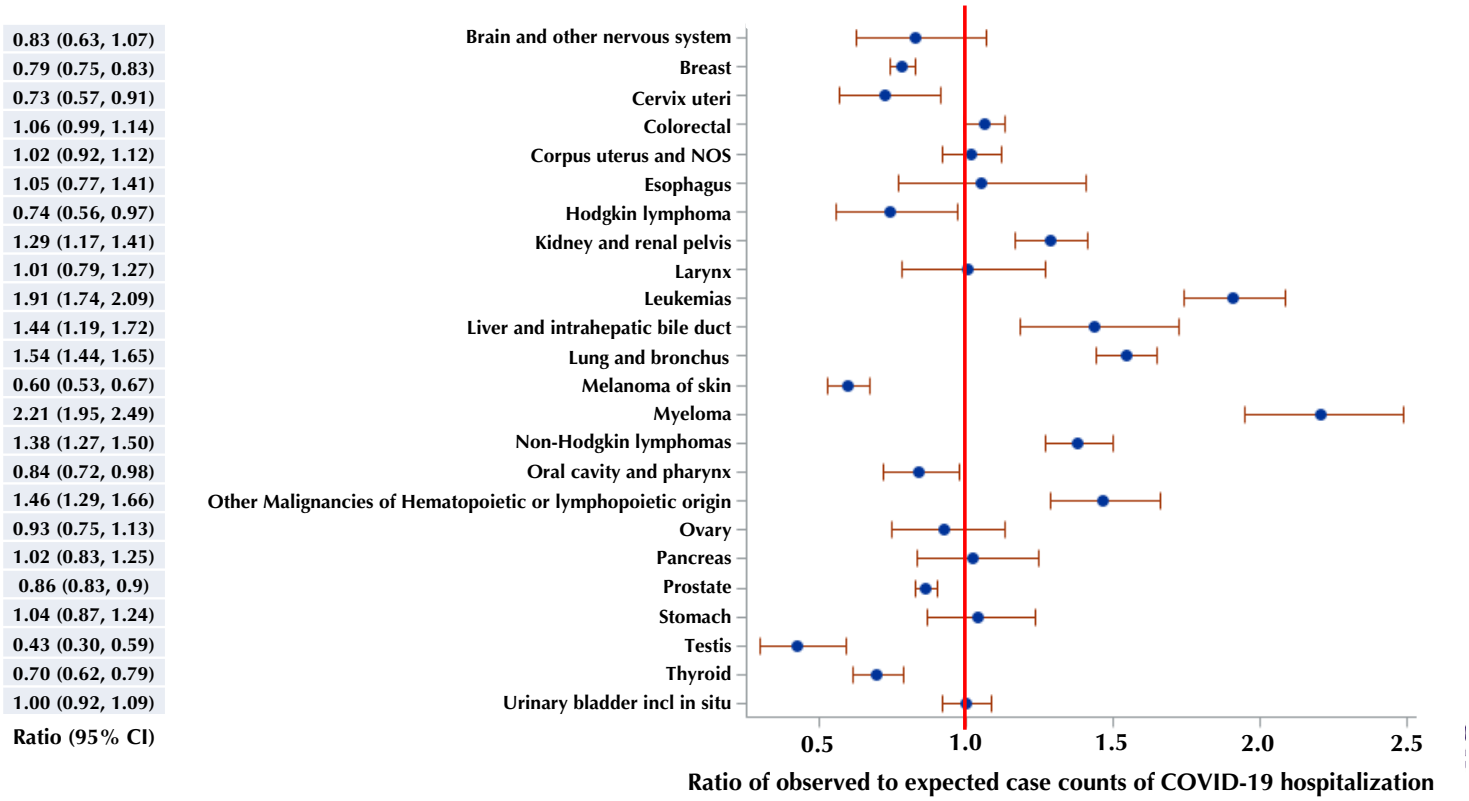
Cancer type	*Number of observed COVID-19 hospitalizations*	*Number of expected COVID-19 hospitalizations**	*Ratio of observed to expected (95% CI)*
Brain and other nervous system	58	70	0.83 (0.63–1.07)
Breast	1,427	1,816	0.79 (0.75–0.83)
Cervix uteri	74	102	0.73 (0.57–0.91)
Colorectal	889	836	1.06 (0.99–1.14)
Corpus uterus and NOS	408	402	1.02 (0.92–1.12)
Esophagus	46	44	1.05 (0.77–1.41)
Hodgkin lymphoma	54	73	0.74 (0.56–0.97)
Kidney and renal pelvis	437	339	1.29 (1.17–1.41)
Larynx	71	71	1.01 (0.79–1.27)
Leukemias	491	257	1.91 (1.74–2.09)
Liver and intrahepatic bile duct	115	80	1.44 (1.19–1.72)
Lung and bronchus	852	552	1.54 (1.44–1.65)
Melanoma	268	448	0.60 (0.53–0.67)
Multiple myeloma	268	121	2.21 (1.95–2.49)
Non-Hodgkin lymphomas	586	424	1.38 (1.27–1.50)
Oral cavity and pharynx	172	204	0.84 (0.72–0.98)
Other malignancies of hematopoietic or lymphopoietic origin	243	166	1.46 (1.29–1.66)
Ovary	94	102	0.93 (0.75–1.13)
Pancreas	100	98	1.02 (0.83–1.25)
Prostate	2,070	2,396	0.86 (0.83–0.90)
Stomach	127	122	1.04 (0.87–1.24)
Testis	36	84	0.43 (0.30–0.59)
Thyroid	257	367	0.70 (0.62–0.79)
Urinary bladder, including in situ	550	549	1.00 (0.92–1.09)

NOS, not otherwise specified. *Number of expected COVID-19 hospitalizations was calculated using the age- and sex-specific numbers of hospitalizations for all cancers combined.

In multivariable-adjusted logistic regression analyses, we observed lower odds of COVID-19 hospitalization among younger age groups and increased odds among older age groups, with the highest odds among individuals ages 60 years and older. Compared to individuals aged 40–49 years, the ORs (95% CIs) were 1.73 (1.52–1.96) for individuals ages 60–69 years, 2.59 (2.28–2.93) for individuals ages 70-79 years, and 4.00 (3.53–4.54) for individuals 80 years of age and older. In addition, we observed increased odds for males versus females (OR, 1.35; 95% CI, 1.29–1.40) and for non-Hispanic Black (OR, 1.75; 95% CI, 1.66–1.85) and Hispanic (OR, 1.66; 95% CI, 1.56–1.77) versus non-Hispanic White individuals. Risk was also increased for individuals who resided in NYC at the time of cancer diagnosis versus the rest of NYS (OR, 1.21; 95% CI, 1.16–1.27); individuals with regional stage (OR, 1.09; 95% CI, 1.03–1.15) or distant stage (OR, 1.69; 95% CI, 1.61–1.79) versus local stage disease; and individuals with multiple cancers versus a single cancer (OR, 1.23; 95% CI, 1.16–1.30). Individuals with more recent cancer diagnosis had higher risk, with increasing odds for decreasing time since cancer diagnosis. Compared to individuals whose most recent cancer was diagnosed more than 10 years in the past, the ORs were 1.53 for cancers diagnosed in the past 0–2 years (95% CI, 1.44–1.62); 1.42 for 3–5 years (95% CI, 1.34–1.50); and 1.25 for 6–10 years (95% CI, 1.19–1.32).

Since region of residency (NYC or rest of NYS) was determined based on address at cancer diagnosis and not address at the time of the COVID-19 pandemic, due to its unavailability for individuals without a linked record in the SPARCS data, we also considered models that did not adjust for region of residence. We obtained very similar results (displayed in column 6 of Table 1) to those obtained from the analysis adjusted for region of residency for all covariates except race/ethnicity. After removing region of residency from the model, the associations with COVID-19

Figure 1. Ratio and 95% CI of Observed to Expected Number of COVID-19 Hospitalizations Among Individuals with a History of Cancer by Cancer Type



hospitalization were strengthened for non-Hispanic Black (OR, 1.91; 95% CI, 1.81–2.01) and Hispanic individuals (OR, 1.83; 95% CI, 1.73–1.94) compared to non-Hispanic White individuals.

Table 2 and Figure 1 display the observed versus expected counts of COVID-19 hospitalizations among individuals with a history of cancer by cancer type, where the expected numbers were determined based on age- and sex-adjusted proportions for all cancers combined. We observed a statistically significant higher than expected number of COVID-19 hospitalizations for several cancers including kidney and renal pelvis cancer (O/E, 1.29; 95% CI, 1.17–1.41), leukemias (O/E, 1.91; 95% CI, 1.74–2.09), liver and intrahepatic bile duct cancer (O/E, 1.44; 95% CI, 1.19–1.72), lung and bronchus cancer (O/E, 1.54; 95% CI, 1.44–1.65), myeloma (O/E, 2.21; 95% CI, 1.95–2.49), non-Hodgkin lymphomas (O/E, 1.38; 95% CI, 1.27–1.50), and other malignancies of hematopoietic or lymphopoietic origin (O/E, 1.46; 95% CI, 1.29–1.66). We observed a statistically significant lower than expected number of COVID-19 hospitalizations for breast cancer (O/E, 0.79; 95% CI, 0.75–0.83), cervix uteri cancer (O/E, 0.73; 95% CI, 0.57–0.91), Hodgkin lymphoma (O/E, 0.74; 95% CI, 0.56–0.97), melanoma (O/E, 0.60; 95% CI, 0.53–0.67), oral cavity and pharynx cancer (O/E, 0.84; 95% CI, 0.72–0.98), prostate cancer (O/E, 0.86; 95% CI, 0.83–0.90), testis cancer (O/E, 0.43; 95% CI, 0.30–0.59), and thyroid cancer (O/E, 0.70; 95% CI, 0.62–0.79).

Table 3 shows the age- and sex-adjusted proportions of death at discharge among all COVID-19 hospitalizations retrieved from SPARCS, by cancer status. The adjusted percentage of individuals with death at discharge was higher among individuals with a prior cancer diagnosis (17.6%) compared to those with no prior cancer diagnosis (15.5%) ($P < .0001$). Among the individuals with a prior cancer diagnosis, we observed a higher adjusted percentage with death at discharge among individuals with a cancer diagnosis claim code at the time of COVID-19 hospitalization (27.1%) compared to those with no cancer diagnosis claim code (20.8%) ($P < .0001$).

Table 4 displays the distribution of different demographic and tumor characteristics of interest and multivariable-adjusted ORs and 95% CIs for each variable in relation to death at discharge among 10,210 individuals with COVID-19 hospitalization and a prior cancer diagnosis. Comparing individuals who were deceased at discharge with those who were not, there were higher percentages of individuals aged 80 years and over, individuals who were male, non-Hispanic Black, non-Hispanic Asian/Pacific Islander (API), Hispanic, living in NYC, diagnosed with distant-stage or unknown-stage cancer or with multiple tumors, or who had active cancer diagnosis claim codes (all $P \leq .01$).

In multivariable-adjusted logistic regression analyses for the associations with death at discharge, we observed increased odds among older age groups, with increasing odds corresponding to increasing age. Compared to individuals 40–49 years of age, we observed ORs (95% CI) of 2.16 (1.41–3.29) for individuals aged 60–69 years, 3.02 (1.99–4.58) for those aged 70–79 years, and 4.85 (3.19–7.38) for those aged 80 years and older. The ORs for younger individuals

Table 3. Crude and Age- and Sex-Adjusted Proportions of Death at Discharge Among 77,334 COVID-19 Hospitalizations for NYS Residents Retrieved from SPARCS, by Cancer Status, and Among 10,210 COVID-19 Hospitalizations Among NYS Residents with a History of Cancer, by Active Status of Cancer

Cancer status	*Deceased at discharge (n)*	*Total N*	*Crude proportion**	*Adjusted proportion***	*SE*	*95% CI*	
No prior cancer diagnosis	9,835	67,124	14.65	15.54	0.16	15.24	15.85
Prior cancer diagnosis	2,328	10,210	22.80	17.62	0.46	16.73	18.52
Inactive cases (no cancer diagnosis claim code)	1,535	7,077	21.69	20.75	0.53	19.70	21.79
Active cases (with cancer diagnosis claim code)	793	3,133	25.31	27.10	0.99	25.16	29.05

NYS, New York State; SPARCS, Statewide Planning and Research Cooperative System.

*χ^2 $P < .0001$ for differences between crude proportions by cancer status and by active status of cancer.

**Adjusted proportions were calculated using SAS STDRATE with the direct standardization method. Age- and sex-specific proportions of deaths for 77,334 COVID-19 hospitalizations were used as the reference in the calculation of adjusted proportions by cancer status, and age- and sex-specific proportions of deaths for 10,210 COVID-19 hospitalizations with prior cancer were used as the reference in the calculation of adjusted proportions by active status of cancer. χ^2 $P < .0001$ comparing individuals with versus without prior diagnosis of cancer and comparing active and inactive cancer diagnoses.

were not statistically significant, and for individuals aged 0–19, the OR was not estimable due to the number of deaths in this age group. In addition, we observed increased odds of death at discharge for males versus females (OR, 1.33; 95% CI, 1.20–1.46); non-Hispanic Black (OR, 1.32; 95% CI, 1.16–1.51), non-Hispanic API (OR, 1.39; 95% CI, 1.11–1.75), and Hispanic individuals (OR, 1.28; 95% CI, 1.11–1.49) compared to non-Hispanic White individuals. In addition, odds of death at discharge were increased for individuals living in NYC at the time of COVID-19 hospitalization (OR, 1.26; 95% CI, 1.13–1.40) compared to the rest of NYS; and individuals with an active cancer diagnosis claim code (OR, 1.39; 95% CI, 1.23–1.58) compared to those with no cancer diagnosis claim code. We did not observe statistically significant associations with death at discharge for stage of cancer or presence of multiple tumors (all $P > .05$). For time since cancer diagnosis, there was a suggestion of decreased odds of death at discharge for individuals diagnosed 6–10 years in the past, compared to those diagnosed more than 10 years previously (OR, 0.87; 95% CI, 0.76–0.99), but, overall, the association with time since cancer diagnosis was not statistically significant ($P = .11$).

Table 5 and Figure 2 show the observed versus expected numbers of deaths at discharge for COVID-19 hospitalization among individuals with a history of cancer by cancer type, where the expected numbers of deaths were determined based on the age- and sex-specific counts for all 10,210 COVID-19 hospitalizations with a history of cancer. We did not observe a higher than expected number of deaths at discharge for any cancer type. However, we observed a lower than expected number of deaths at discharge for breast cancer (O/E, 0.73; 95% CI, 0.64–0.82), corpus uterus and NOS (O/E, 0.76; 95% CI, 0.60–0.96), melanoma (O/E, 0.64; 95% CI, 0.47–0.85), and thyroid cancer (O/E, 0.67; 95% CI, 0.47–0.93), after adjustment for age and sex.

Discussion

In this population-based analysis, we linked data on all NYS residents who had a history of invasive cancer and were alive immediately prior to the start of the COVID-19 pandemic with claims data on hospitalizations for COVID-19 in 2020 and the first half of 2021. We observed that individuals with a history of cancer were 16% more likely to be hospitalized for COVID-19, compared to the general population of NYS. Factors independently associated with COVID-19 hospitalization among cancer patients included older age, male sex, non-Hispanic Black race or Hispanic ethnicity, diagnosis with late-stage cancer or with multiple tumors, more recent cancer diagnosis, and NYC residency at the time of cancer diagnosis. In addition, we observed that individuals with a history of cancer were more likely to die while hospitalized for COVID-19, compared to those with no prior cancer diagnosis, and among individuals with a history of cancer, those with a cancer diagnosis claim code (indicating an active cancer diagnosis) were more likely to die than those without a cancer diagnosis claim code. Factors independently associated with death at discharge among individuals with a prior cancer diagnosis included older age, male sex, non-Hispanic Black or non-Hispanic API race or Hispanic ethnicity, living in NYC at the time of COVID-19 hospitalization, and having an active cancer diagnosis claim code.

The results of this study indicate that individuals with a history of cancer are at an increased risk for severe COVID-19 outcomes, including hospitalization and death, which is in agreement with findings from previous studies.[5-12] Consistent with prior studies, we also observed that older age,[5,11,12] Black race,[7,20] diagnosis with late-stage cancer,[16] and more recent cancer diagnosis[7,16] were independently associated with severe COVID-19 among individuals with a history of cancer. However, to our knowledge, our findings that diagnosis with multiple tumors and residence in NYC were associated with COVID-19 hospitalization in

Table 4. Distribution of Demographic Variables and Tumor Characteristics of Interest and Multivariable-Adjusted ORs and 95% CIs for Each Variable and Vital Status at Discharge Among 10,210 Individuals with COVID-19 Hospitalization and Prior Cancer Diagnosis

Variable	*Cancer cases deceased at discharge from COVID-19 hospitalization (n = 2,328), n (%)*	*Cancer cases alive at discharge from COVID-19 hospitalization (n = 7,882), n (%)*	*P value**	*OR (95% CI)***
Age in years, mean (SD)	77.9 (10.9)	72.8 (13.3)	<.0001	
Age group (y)			<.0001	
0–19	<11	30 (0.4)		NA
20–29	<11	37 (0.5)		0.24 (0.03–1.81)
30–39	<11	69 (0.9)		0.26 (0.06–1.13)
40–49	26 (1.1)	241 (3.1)		Ref
50–59	99 (4.3)	742 (9.4)		1.29 (0.82–2.04)
60–69	365 (15.7)	1,676 (21.3)		2.16 (1.41–3.29)
70–79	726 (31.2)	2,523 (32.0)		3.02 (1.99–4.58)
≥80	1,109 (47.6)	2,564 (32.5)		4.85 (3.19–7.38)
Sex			<.0001	
Male	1,415 (60.8)	4,212 (53.4)		1.33 (1.20–1.46)
Female	913 (39.2)	3,670 (46.6)		Ref
Race/ethnicity			0.001	
Non-Hispanic White	1,352 (58.1)	4,924 (62.5)		Ref
Non-Hispanic Black	482 (20.7)	1,448 (18.4)		1.32 (1.16–1.51)
Non-Hispanic API	121 (5.2)	333 (4.2)		1.39 (1.11–1.75)
Hispanic	356 (15.3)	1,101 (14.0)		1.28 (1.11–1.49)
Non-Hispanic other/unknown	17 (0.7)	76 (1.0)		0.85 (0.50–1.47)
Region of residency***			<.0001	
New York City	1,159 (49.8)	3,359 (42.6)		1.26 (1.13–1.40)
Rest of New York State	1,169 (50.2)	4,523 (57.4)		Ref
Stage of Cancer			0.005	
Local	1,072 (46.1)	3,854 (48.9)		Ref
Regional	354 (15.2)	1,274 (16.2)		1.07 (0.92–1.23)
Distant	474 (20.4)	1,509 (19.1)		1.12 (0.98–1.30)
Unknown stage	428 (18.4)	1,245 (15.8)		1.12 (0.98–1.29)
Multiple Tumor			0.012	
No	1,901 (81.7)	6,611 (83.9)		Ref
Yes	427 (18.3)	1,271 (16.1)		1.03 (0.91–1.18)
Time since cancer diagnosis			0.003	
0–2 years (2019–2021)	411 (17.7)	1,429 (18.1)		0.96 (0.83–1.13)
3–5 years (2016–2019)	521 (22.4)	1,764 (22.4)		1.02 (0.89–1.17)
6–10 years (2012–2016)	492 (21.1)	1,912 (24.3)		0.87 (0.76–0.99)
>10 years (1976–2011)	904 (38.8)	2,777 (35.2)		Ref
Active cancer diagnosis			<.0001	
Cancer diagnosis claim code	793 (34.1)	2,340 (29.7)		1.39 (1.23–1.58)
No cancer diagnosis claim code	1,535 (65.9)	5,542 (70.3)		Ref

*P-values from χ^2 tests for categorical variables and Student's t test for mean age.
**Analyses mutually adjusted for all variables in the table.
***Region of residency determined based on patient address at COVID-19 hospitalization from data obtained from SPARCS.

Table 5. Ratio and 95% CI of Observed to Expected Number of Deaths at Discharge Among 10,210 Individuals with COVID-19 Hospitalization and a History of Cancer, by Cancer Type

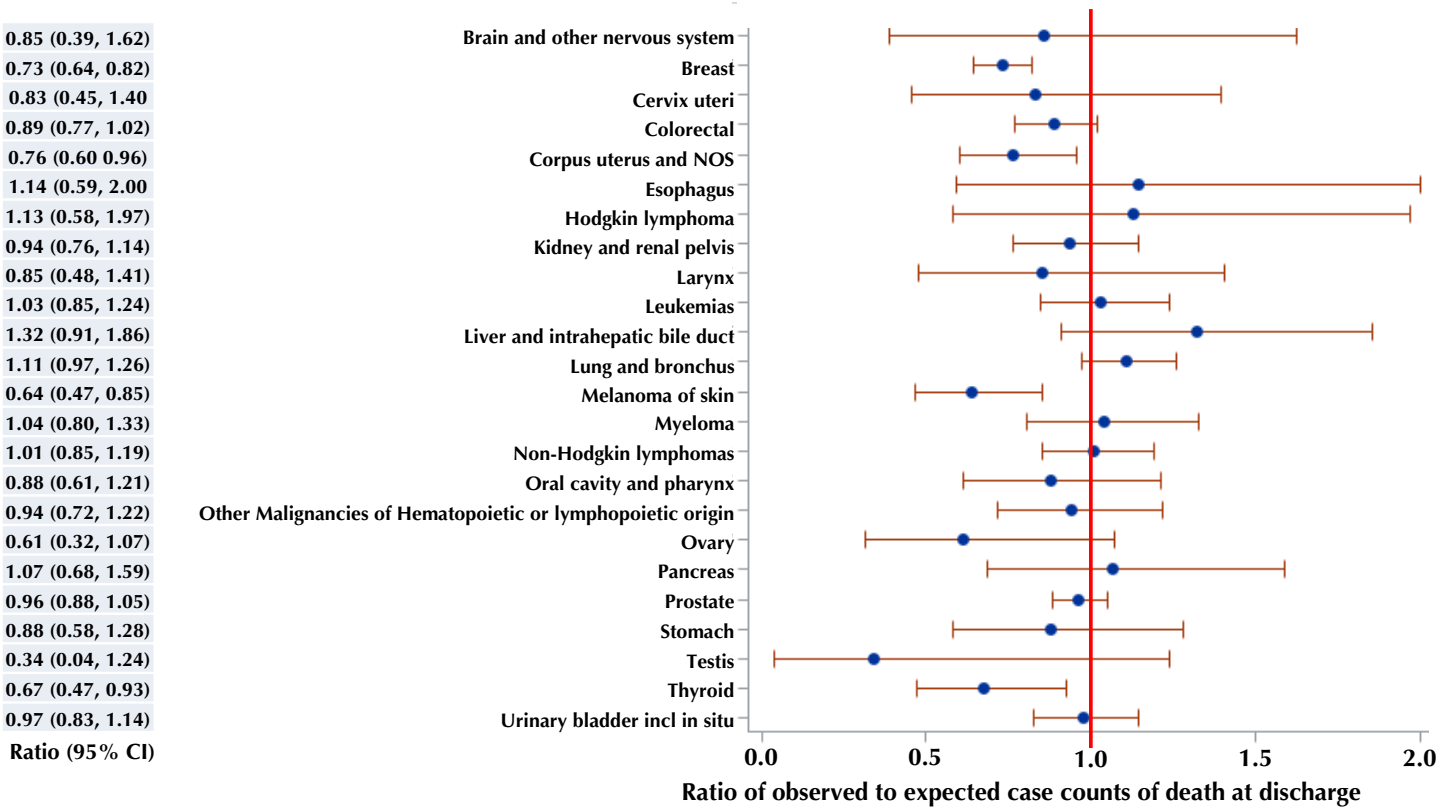
Cancer type	*Number of observed deaths*	*Number of expected deaths**	*Ratio of observed to expected (95% CI)*
Brain and other nervous system	<11	Suppressed	0.85 (0.39–1.62)
Breast	267	366	0.73 (0.64–0.82)
Cervix uteri	14	17	0.83 (0.45–1.40)
Colorectal	207	233	0.89 (0.77–1.02)
Corpus uterus and NOS	75	98	0.76 (0.60–0.96)
Esophagus	12	10	1.14 (0.59–2.00)
Hodgkin lymphoma	12	11	1.13 (0.58–1.97)
Kidney and renal pelvis	98	105	0.94 (0.76–1.14)
Larynx	15	18	0.85 (0.48–1.41)
Leukemias	113	110	1.03 (0.85–1.24)
Liver and intrahepatic bile duct	33	25	1.32 (0.91–1.86)
Lung and bronchus	235	212	1.11 (0.97–1.26)
Melanoma	46	72	0.64 (0.47–0.85)
Myeloma	65	62	1.04 (0.80–1.33)
Non-Hodgkin lymphomas	140	138	1.01 (0.85–1.19)
Oral cavity and pharynx	36	41	0.88 (0.61–1.21)
Other malignancies of hematopoietic or lymphopoietic origin	58	61	0.94 (0.72–1.22)
Ovary	12	20	0.61 (0.32–1.07)
Pancreas	24	22	1.07 (0.68–1.59)
Prostate	532	552	0.96 (0.88–1.05)
Stomach	27	31	0.88 (0.58–1.28)
Testis	<11	Suppressed	0.34 (0.04–1.24)
Thyroid	37	55	0.67 (0.47–0.93)
Urinary bladder including in situ	153	157	0.97 (0.83–1.14)

NOS, not otherwise specified. * Number of expected COVID-19 hospitalizations was calculated using the age- and sex-specific proportions of deaths for all 10,210 cancers.

individuals with a history of cancer have not been identified in previous studies. Although the association with residence in NYC was based on address at the time of cancer diagnosis, and not at the time of COVID-19 hospitalization, this variable likely approximated residence at the time of hospitalization for a majority of patients. NYC was an epicenter of the COVID-19 pandemic and experienced a large number of COVID-19 cases, hospitalizations, and deaths early on in the pandemic, which likely contributed to this finding of an increased risk of hospitalization compared to residents of the rest of NYS.[21,22] We were unable to assess differences in the total number of COVID-19 cases among cancer patients by variables such as region of NYS, race/ethnicity, and sex. However, it is likely that at least part of the difference in the risk of COVID-19 hospitalization by demographic characteristics was related to differences in the overall number of COVID-19 cases across population groups.[21,22] A higher risk of severe COVID-19 in certain demographic groups, regardless of cancer status, may have also contributed to the results.[22,23]

In adjusted analyses of the observed versus expected number of hospitalizations by cancer type, we observed a higher than expected number of COVID-19 hospitalizations for kidney and renal pelvis cancer, leukemias, liver and intrahepatic bile duct cancer, lung and bronchus cancer, multiple myeloma, non-Hodgkin lymphomas, and other malignancies of hematopoietic or lymphopoietic origin. Previous studies have reported similar results for liver,[16] lung,[16] and hematological malignancies,[16,24] but to our knowledge no prior studies have examined as many as 24 cancer types, and we observed associations with certain cancer types that were not previously reported. We did not have data on the presence of comorbidities, including HIV, that are more common in patients with certain cancers and that may increase the risk of hospitalization after diagnosis with COVID-19.[4,16,25] It is possible that some of the increased risk of COVID-19 hospitalization for certain cancer types was related to the presence of HIV, which is associated with increased risk of several cancers including non-Hodgkin lymphoma, lung cancer, and liver cancer,[26] or other comorbidities.

Figure 2. Ratio and 95% CI of Observed to Expected Number of Deaths at Discharge Among 10,210 Individuals with COVID-19 Hospitalization and a History of Cancer by Cancer Type



We observed that older age, male sex, non-Hispanic Black or non-Hispanic API race or Hispanic ethnicity, living in NYC at the time of COVID-19 hospitalization, and having an active cancer diagnosis claim code were associated with higher risk of death at discharge among 10,210 COVID-19 hospitalizations of patients with a history of cancer. Our findings were consistent with other studies that observed increased mortality with older age,[10,11,12,16] male sex,[10,11] and Black race.[16,20] Our analyses did not indicate that diagnosis with late-stage cancer or multiple tumors or more recent cancer diagnosis was associated with higher risk of death at discharge. One previous study reported that the presence of multisite tumors was associated with increased risk of all-cause mortality, although the paper noted that this included any patient where more than 1 cancer site was reported and that cases with subsequent malignancy versus metastasis could not be distinguished.[10] To our knowledge, no previous studies have examined associations between residence in NYC and mortality at COVID-19 discharge among patients with a history of cancer. This association may again be related to the severity of the early COVID-19 pandemic in NYC, compared to the rest of NYS. The association between an active cancer diagnosis claim code and higher risk of death at discharge is likely due to immunosuppression related to the cancer itself or treatment, or other effects of recent cancer treatment. Consistent with our findings, a prior study observed that patients with COVID-19 and recent cancer treatment had a higher risk of death (OR, 1.74; 95% CI ,1.54–1.96), while those with no recent cancer treatment did not have increased mortality (OR, 0.93; 95% CI, 0.84–1.02), compared to individuals without cancer.[15] Some of the associations between demographic characteristics and death at discharge may be related to poorer outcomes in cancer patients with these characteristics, regardless of COVID-19 status. However, the occurrence of death during hospitalization for COVID-19 and the strength of the associations suggest a clear role of COVID-19 in these outcomes.

In adjusted analyses of the observed versus expected number of deaths by cancer type, we did not observe a statistically significant increased risk of death at discharge for any cancer type. In contrast, some previous studies reported that hematologic malignancies[8,9,10,24] and lung cancer[10] were associated with an increased risk of mortality after COVID-19. Possible reasons for this inconsistency may be differences in the cancer case selection or comparison group, as we only analyzed the death at discharge among patients with COVID-19-related hospitalizations, rather than all COVID-19 patients, and compared the mortality for each cancer type to that for all cancers combined. The use of a different comparison group, such as all COVID-19-related hospitalizations, would be expected to yield lower counts of the expected number of deaths and higher ratios of the observed to expected number of deaths. However, patients with a history of cancer may differ from other patients hospitalized for COVID-19 in ways that could not be controlled for in this analysis but that would be expected to impact mortality, such as having a higher number of comorbidities.[9,15,16]

Our results indicate an increased risk of COVID-19 hospitalization and death among cancer patients, and

in particular those with certain demographic and tumor characteristics. The explanation for these findings is likely multifactorial and related to both immune function and risk patterns of severe COVID-19. A poorer immune response or worse course of illness in cancer patients that is related to the cancer itself, cancer treatment, or a higher prevalence of cancer-related comorbidities would be expected to lead to poorer outcomes after diagnosis with COVID-19. This is supported by the associations we observed with late-stage disease, multiple tumors, more recent diagnosis, and active cancer. Severity of COVID-19 due to diagnosis early in the pandemic, prior to the availability of effective treatments and vaccines, or reduced access to care likely also contributed to some of the results we observed, including the stronger associations for NYC, an early epicenter of the pandemic where the impact was greatest on vulnerable populations.[22] Other associations may be related to both immune-related factors and COVID-19 severity, including the stronger associations observed among older individuals, men, and certain racial and ethnic groups. These demographic characteristics have been associated with more severe COVID-19 regardless of cancer status,[27-29] and the associations we observed likely have both cancer-related and independent contributing factors including immune function, comorbidity burden, and patterns of care.

Strengths of this study include the availability of statewide population-based data on cancer diagnoses and COVID-19 hospitalizations, including patient demographics and case characteristics. This allowed us to look at a number of predictors of COVID-19 hospitalization and death at discharge in cancer patients. However, a small proportion of COVID-19 hospitalizations were likely missed, including those that occurred at Veterans Affairs and other military hospitals that are not captured in SPARCS. In addition, the match of the NYSCR and SPARCS data may have missed some cases that were not identified by our deterministic matching methods, which would be expected to result in an underestimation of the risk of COVID-19 hospitalization for cancer patients. For patients with multiple tumors, we used the type, stage, and date of diagnosis for the most recent tumor; however, in some cases this may not be the most relevant cancer diagnosis for the patient's health. Additionally, for analyses of risk of COVID-19 hospitalization among all patients with a history of cancer, region of NYS was categorized based on address at diagnosis for the most recent tumor and may not reflect a patient's current address, particularly for cases diagnosed further in the past. Finally, by using the first COVID-19 hospitalization record for patients who had multiple hospitalizations, we may have undercounted deaths among all COVID-19 patients or among COVID-19 patients with a prior cancer. Although we did not examine subsequent hospitalizations, a previous study of US electronic health record and administrative data reported that during the period from March to August 2020, 9% of patients hospitalized with COVID-19 were readmitted to the same hospital within 2 months of discharge but less than 0.1% of patients died during readmission, suggesting that only a small number of deaths were missed by omitting subsequent hospitalizations.[30]

In summary, our results indicate that cancer patients were more likely to be hospitalized for COVID-19 than individuals without a history of cancer, and, among cancer patients, several case characteristics and cancer types were associated with an increased risk of COVID-19 hospitalization. In addition, patients with a history of cancer had a statistically significant increased risk of death after COVID-19 hospitalization, compared to patients without a history of cancer, and this risk was strongest among certain demographic groups and patients with an active cancer claim code at the time of their COVID-19 hospitalization. Consistent with the results of most previous studies, our results indicate a higher risk of severe COVID-19 outcomes among cancer patients, and in particular those in certain demographic groups or with certain diagnostic characteristics. Although this study focused on hospitalizations and deaths during the early part of the COVID-19 pandemic, prior to the widespread availability of vaccines and treatments for COVID-19, the results highlight the importance of continued vigilance to ensure the best possible outcomes for all patients with a history of cancer.

References

1. *Cancer Treatment & Survivorship Facts & Figures 2022–2024.* American Cancer Society; 2022.
2. Graphs. National Cancer Institute Division of Cancer Control and Population Sciences website. Accessed September 25, 2022. **https://cancercontrol.cancer.gov/ocs/statistics#graphs**
3. Centers for Disease Control and Prevention. COVID-19: underlying medical conditions associated with higher risk for severe COVID-19: information for healthcare professionals. Updated June 15, 2022. **https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html**
4. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. Science brief: evidence used to update the list of underlying medical conditions associated with higher risk for severe COVID-19. In: *CDC COVID-19 Science Briefs* [Internet]. Centers for Disease Control and Prevention; 2020.
5. Robilotti EV, Babady NE, Mead PA, et al. Determinants of COVID-19 disease severity in patients with cancer. *Nat Med*. 2020;26(8):1218-1223. doi:10.1038/s41591-020-0979-0
6. Yang L, Chai P, Yu J, Fan X. Effects of cancer on patients with COVID-19: a systematic review and meta-analysis of 63,019 participants. *Cancer Biol Med*. 2021;18(1):298-307. doi:10.20892/j.issn.2095-3941.2020.0559
7. Wang Q, Berger NA, Xu R. Analyses of risk, racial disparity, and outcomes among US patients with cancer and COVID-19 infection. *JAMA Oncol*. 2021;7(2):220-227. doi:10.1001/jamaoncol.2020.6178
8. Meng Y, Lu W, Guo E, et al. Cancer history is an independent risk factor for mortality in hospitalized COVID-19 patients: a propensity score-matched analysis. *J Hematol Oncol*. 2020;13(1):75. doi:10.1186/s13045-020-00907-0
9. Venkatesulu BP, Chandrasekar VT, Girdhar P, et al. A systematic review and meta-analysis of cancer patients affected by a novel coronavirus. *medRxiv*. Preprinted posted online May 29, 2020. doi:10.1101/2020.05.27.20115303
10. Sharafeldin N, Bates B, Song Q, et al. Outcomes of COVID-19 in patients with cancer: report from the National COVID Cohort Collaborative (N3C). *J Clin Oncol*. 2021;39(20):2232-2246. doi:10.1200/JCO.21.01074
11. Han S, Zhuang Q, Chiang J, et al. Impact of cancer diagnoses on the outcomes of patients with COVID-19: a systematic review and meta-analysis. *BMJ Open*. 2022;12(2):e044661. doi:10.1136/bmjopen-2020-044661
12. Fillmore NR, La J, Szalat RE, et al. Prevalence and outcome of COVID-19 infection in cancer patients: a National Veterans Affairs study. *J Natl Cancer Inst*. 2021;113(6):691-698. doi:10.1093/jnci/djaa159

13. Joharatnam-Hogan N, Hochhauser D, Shiu KK, et al. Outcomes of the 2019 novel coronavirus in patients with or without a history of cancer: a multi-centre North London experience. *Ther Adv Med Oncol*. 2020;12:1758835920956803. doi:10.1177/1758835920956803
14. Klein IA, Rosenberg SM, Reynolds KL, et al. Impact of cancer history on outcomes among hospitalized patients with COVID-19. *Oncologist*. 2021;26(8):685-693. doi:10.1002/onco.13794
15. Chavez-MacGregor M, Lei X, Zhao H, Scheet P, Giordano SH. Evaluation of COVID-19 mortality and adverse outcomes in US patients with or without cancer. *JAMA Oncol*. 2022;8(1):69-78. doi:10.1001/jamaoncol.2021.5148
16. Kim Y, Zhu L, Zhu H, et al. Characterizing cancer and COVID-19 outcomes using electronic health records. *PLoS One*. 2022;17(5):e0267584. doi:10.1371/journal.pone.0267584
17. Statewide Planning and Research Cooperative System (SPARCS). New York State Department of Health website. Revised August 2022. **https://www.health.ny.gov/statistics/sparcs/**
18. Centers for Medicare and Medicaid Services. Preliminary Medicare COVID-19 Data Snapshot Methodology. **https://www.cms.gov/files/document/medicare-covid-19-data-snapshot-methodology.pdf**
19. Breslow NE, Day NE. *Statistical Methods in Cancer Research: Vol. II, The Design and Analysis of Cohort Studies*. IARC Scientific Publication No. 82. International Agency for Research on Cancer; 1987.
20. Fu J, Reid SA, French B, et al. Racial disparities in COVID-19 outcomes among black and white patients with cancer. *JAMA Netw Open*. 2022;5(3):e224304. doi:10.1001/jamanetworkopen.2022.4304
21. Bialek S, Bowen V, Chow N, et al. CDC COVID-19 Response Team. Geographic differences in COVID-19 cases, deaths, and incidence—United States, February 12–April 7, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(15):465-471. doi:10.15585/mmwr.mm6915e4
22. Thompson CN, Baumgartner J, Pichardo C, et al. COVID-19 Outbreak—New York City, February 29–June 1, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(46):1725-1729. doi:10.15585/mmwr.mm6946a2
23. Gu T, Mack JA, Salvatore M, et al. Characteristics associated with racial/ethnic disparities in COVID-19 outcomes in an academic health care system. *JAMA Netw Open*. 2020;3(10):e2025197. doi: 10.1001/jamanetworkopen.2020.25197
24. Wang Q, Berger NA, Xu R. When hematologic malignancies meet COVID-19 in the United States: infections, death and disparities. *Blood Rev*. 2021;47:100775. doi:10.1016/j.blre.2020.100775
25. Tesoriero JM, Swain CE, Pierce JL, et al. COVID-19 outcomes among persons living with or without diagnosed HIV infection in New York State. *JAMA Netw Open*. 2021;4(2):e2037069. doi:10.1001/jamanetworkopen.2020.37069
26. Shiels MS, Islam JY, Rosenberg PS, et al. Projected cancer incidence rates and burden of incident cancer cases in HIV-infected adults in the United States through 2030. *Ann Intern Med*. 2018;168(12):866-873. doi:10.7326/M17-2499
27. St Sauver JL, Lopes GS, Rocca WA, et al. Factors associated with severe COVID-19 infection among persons of different ages living in a defined Midwestern US population. *Mayo Clin Proc*. 2021;96(10):2528-2539. doi:10.1016/j.mayocp.2021.06.023
28. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020;584(7821):430-436. doi:10.1038/s41586-020-2521-4
29. Pijls BG, Jolani S, Atherley A, et al. Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ Open*. 2021;11(1):e044640. doi:10.1136/bmjopen-2020-044640
30. Lavery AM, Preston LE, Ko JY, et al. Characteristics of hospitalized COVID-19 patients discharged and experiencing same-hospital readmission—United States, March–August 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:1695-1699. doi:10.15585/mmwr.mm6945e2

Original Article

Improving Precision of Do Not Contact Codes: Results of a Manual Review to Inform Coding and Case Contact Procedures

Katherine A. Lawson-Michod[a]; Marjorie Carter[b]; Valerie Yoder[b]; Rachel D. McCarty[a]; Carrie Bateman[b]; Morgan M. Millar[a,b,c]; Jennifer A. Doherty[a,b]

Abstract: Introduction: Central cancer registries are responsible for managing appropriate research contacts and record releases. Do not contact (DNC) flags are used by some registries to indicate patients who should not be contacted or included in research. Longitudinal changes in DNC coding practices and definitions may result in a lack of code standardization and inaccurately include or exclude individuals from research. Purpose: We performed a comprehensive manual review of DNC cases in the Utah Cancer Registry to inform updates to standardization of DNC code definitions, and use of DNC codes for exclusion/inclusion in research. Methods: We identified 858 cases with a current or prior DNC flag in the SEER Data Management System (SEER*DMS) or a research database, with cancers diagnosed from 1957–2021. We reviewed scanned images of correspondence with cases and physicians, incident forms, and comments in SEER*DMS and research databases. We evaluated whether there was evidence to support the current DNC code, a different DNC code, or insufficient evidence for any code. Results: Of the 755 cases that had a current DNC flag and reason code in SEER*DMS, the distribution was as follows: 58%, *Patient requested no contact;* 20%, *Physician denied;* 13%, *Patient is not aware they have cancer;* 4%, *Patient is mentally disabled* [*sic*]; 4%, *Other;* and 1%, *Unknown*. In 5% of these cases, we found evidence supporting a different DNC reason code. Among cases included because of a prior DNC flag in SEER*DMS (n = 10) or a DNC flag in a research database (ie, cases with no current DNC flag or reason code in SEER*DMS, n = 93), we found evidence supporting the addition of a SEER*DMS DNC flag and reason code in 50% and 40% of cases, respectively. We identified DNC reason codes with outdated terminology *(Patient is mentally disabled)* and codes that may not accurately reflect patient research preferences (*Physician denied* without asking the patient). To address this, we identified new reason codes, retired old reason codes, and updated current reason code definitions and research handlings. Conclusion: The time and resource investment in manual review allowed us to identify and, in most cases, resolve discordance in DNC flags and reason codes, adding reason codes when they were missing. This process was valuable because it informed recommended changes to DNC code definitions and research handlings that will ensure more appropriate inclusion and exclusion of cancer cases in research.

Key words: *cancer registries, cancer research, do not contact record releases, research contacts, Surveillance, Epidemiology, and End Results (SEER) Program*

Introduction

Central cancer registries are the foundational source of information for cancer surveillance and control in the United States[1,2] and are widely used for population-based research.[3,4] Central registries participating in the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program or the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) are required to maintain a population-based reporting system with annual follow-up that meets specific quality standards.[1,2,5,6] These practices allow for more accurate data reporting and timely dissemination of information for cancer surveillance and control.[3,7,8] Cancer is a reportable disease, and individuals diagnosed with cancer cannot opt out of state-legislated public health surveillance reporting.[9] However, many registries are also engaged in contacting and recruiting representative samples of individuals with cancer or survivors for research, which requires careful tracking and documentation of individuals' preferences for being contacted and having their records released to ensure patient privacy.[1]

The handling of research contacts and record releases differs across registries, including whether physician permission is required before contacting a patient and whether the registry or the researchers make the first contact with the patient.[10] Additionally, some registries are "opt in," meaning a patient must agree to research contacts or record releases, while other registries are "opt out." Many cancer

[a]Huntsman Cancer Institute, Department of Population Health Sciences, University of Utah, Salt Lake City, Utah. [b]Utah Cancer Registry, University of Utah, Salt Lake City, Utah. [c]Department of Internal Medicine, University of Utah, Salt Lake City, Utah.

Address correspondence to Katherine A. Lawson-Michod, 2000 Circle of Hope, Room 4702, Salt Lake City, UT 84112. Telephone: (801) 646-4077. Email: Katherine.lawson-michod@hci.utah.edu

The Utah Cancer Registry is funded by the National Cancer Institute's SEER Program, Contract No. HHSN261201800016I, and the US Centers for Disease Control and Prevention's National Program of Cancer Registries, Cooperative Agreement No. NU58DP007131, with additional support from the University of Utah and Huntsman Cancer Foundation. The research reported in this publication was also supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002538**.** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

registries have developed tools for managing and documenting patient contacts and record releases. For example, in the SEER Data Management System (SEER*DMS), a patient record can be flagged as *Do not contact* (DNC), and a reason code explaining the flag can be provided.[11] Because of different processes for handling research contacts and recorded releases, the use and interpretation of the DNC flag and reason code are not standardized across registries. Additionally, coding practices and code definitions can change over time within a registry. Although this flexibility allows registries to use the DNC features to meet their current operational needs, changes in code definitions over time may inaccurately exclude individuals from opportunities to participate in studies requiring contact or research linkages.

The Utah Cancer Registry is a SEER and NPCR registry that receives reports for approximately 14,000 new cancer diagnoses every year.[12] Cancer was first designated as a reportable disease in Utah in 1948. However, systematic cancer surveillance was not conducted in the state until 1966. The Utah Cancer Registry was designated as a SEER site in 1973 and joined NPCR in 2017. The Utah Cancer Registry routinely contacts cancer patients or survivors for research, conducts linkages for research studies, and generates research data sets for investigators. The Utah Cancer Registry uses the DNC flag in SEER*DMS to indicate individuals who should not be contacted and whose records potentially should be excluded from research linkages. DNC reason codes provide the rationale for a flag and can be assigned at either the tumor or patient level. Tumor-level flags should exclude individuals from research contacts for that diagnosis only (eg, if the person is not aware of the cancer diagnosis), while patient-level flags exclude individuals from research contacts for all diagnoses and specific database linkages. Contact events with individuals or health care providers can result in a DNC flag. DNC reason codes correspond to situations arising from common patient responses during a contact event, such as, "I do not want to be contacted" or "I do not have cancer." Additionally, research databases for specific projects have a separate DNC data field used by research staff. Research databases do not automatically transfer DNC codes to SEER*DMS, and manual transfers from these databases may have missed, incorrectly assigned, or even erased codes. Additionally, while the Utah Cancer Registry DNC flag is now used only for research contacts and linkages, it was historically used for operational practices such as contact events for follow-up. Concern that these longitudinal changes may have resulted in a lack of code standardization and the realization that contemporizing current code definitions and practices were necessary prompted a manual review of records in the Utah Cancer Registry. We reviewed records with a current or prior DNC flag in SEER*DMS or a DNC code in our research database for evidence supporting the flag and reason code. This manual review informed efforts to appropriately update and standardize codes and processes for coding DNCs and handling research contacts within the Utah Cancer Registry.

Methods

Identification of Cancer Cases with a DNC Flag

We sought to perform a comprehensive review of all cancer cases in the Utah Cancer Registry that were flagged as DNC either in SEER*DMS or in Utah Cancer Registry research databases. Of 401,382 cases in the Utah Cancer Registry diagnosed between 1948 and 2020, we identified 765 with a current (n = 755) or prior (n = 10) DNC flag in SEER*DMS. In order to identify cases that may have been miscoded in SEER*DMS or for which the SEER*DMS code was erased, we also searched Utah Cancer Registry research databases to identify cases that did not have a DNC flag in SEER*DMS, but did have a flag indicating DNC in the research database. Using this method, we identified another 93 cases, for a total of 858 cases with a DNC flag. These cases were diagnosed between 1957 and 2020.

Manual Review of Documentation of DNC Flags and Reason Codes

For all 858 cases, we reviewed all available documentation for evidence related to a current DNC flag or reason code. Documentation included scanned images of correspondence with individuals and physicians, including incident forms, comments in SEER*DMS, and codes and comments in linked research databases. Some sources were from historic practices that have been discontinued, including patient responses to annual contact letters previously used for follow-up and permission request forms sent to physicians. We evaluated whether there was evidence to support the current DNC code, evidence to support a different DNC code, or insufficient evidence for any code. We recorded the evidence type, location, date, description, and the recommended code based on the evidence found. If a record had evidence to support the historic *Physician denied* reason code, the reason for denial was recorded. For cases with evidence supporting multiple codes, the evidence for all codes was recorded, but a single patient-level code with the highest levels of restrictions for case contact and record release was recommended. A trained research assistant completed the manual review and recorded results in a REDCap database. Training included an introduction to SEER*DMS and research databases as well as a pilot review of 20 cases led by Utah Cancer Registry managers. Throughout the manual review, the reviewer met with the Utah Cancer Registry managers to provide updates on the results of the manual review and discuss coding recommendations, conflicting evidence, and any other questions. We defined *discordant* cases as those with an absence of evidence for any code or evidence that did not support the current DNC flag or reason code.

Results

All reviewed flags and codes were at the patient, not tumor, level. Of the 858 reviewed cases, most (88%) had a current DNC flag in SEER*DMS, though 1% had a prior DNC flag in SEER*DMS, and another 11% had a DNC flag in a research database but no DNC flag in SEER*DMS (Table 1). Among the 755 cases with a current SEER*DMS DNC

Table 1. Distribution of Cases Included in the Manual Review

Source of cases included in the manual review	n	%
Current SEER*DMS DNC flag	755	88
DNC in research database	93	11
Prior DNC in SEER*DMS	10	1
Total	858	100
Distribution of reason codes among cases with a current DNC flag in SEER*DMS	**n**	**%**
Patient requested no contact (includes requests to have records removed from registry)	440	58
Physician denied	150	20
Patient is not aware they have cancer	100	13
Patient is mentally disabled [sic]	27	4
Other	27	4
Unknown	11	1
Total	755	100

DMS, Data Management System; DNC, do not contact.

Cases were identified through 3 sources: those with a current DNC flag in SEER*DMS; those with a prior DNC flag in SEER*DMS; and those with a DNC flag in a research database. Only cases with a current DNC flag in SEER*DMS had a corresponding reason code.

flag, the distribution of the current accompanying DNC reason codes was as follows: *Patient requested no contact* (n = 440; 58%); *Physician denied* (n = 150; 20%); *Patient is not aware they have cancer* (n = 100; 13%); *Patient is mentally disabled* [*sic*] (n = 27; 4%); *Other* (n = 27; 4%); *and Unknown* (n = 11; 1%) (Table 1).

Among the 755 cases with a current SEER*DMS DNC flag, we found evidence supporting the current SEER*DMS code status (ie, supporting the presence of a flag and specific reason code) for 402 cases (53%), evidence supporting a different DNC reason code in 38 cases (5%), and no evidence to support either a flag or a code in 315 cases (42%). Of the 402 cases with evidence supporting the current SEER*DMS code status, 32 (8%) had evidence supporting a different code in addition to the current code. All cases identified in this category had evidence supporting a combination of 2 of the following reason codes: *Patient requested their data be removed from registry*, *Patient requested no contact*, or *Patient is not aware they have cancer*. Of the 315 with no evidence to support either a flag or a code, 150 currently have a DNC flag and historic reason code of *Physician denied*. This code was based on a historic form sent to physicians that asked them for permission to "locate" the patient for follow-up, not for permission to contact the patient for research purposes. We therefore did not consider responses to this physician contact form to be evidence. The other 165 cases with no evidence to support either a flag or a code had no documentation of any type in SEER*DMS.

Among cases with no current DNC flag or reason code in SEER*DMS that were included because of a prior DNC flag in SEER*DMS (n=10) or a DNC flag in a research database (n=93), we found evidence supporting the addition of a SEER*DMS DNC flag and reason code for 5 (50%) and 37 (40%) cases, respectively. For the remaining 5 (50%) and 56 (60%) cases in each group, respectively, we found no evidence to support either a flag or a code.

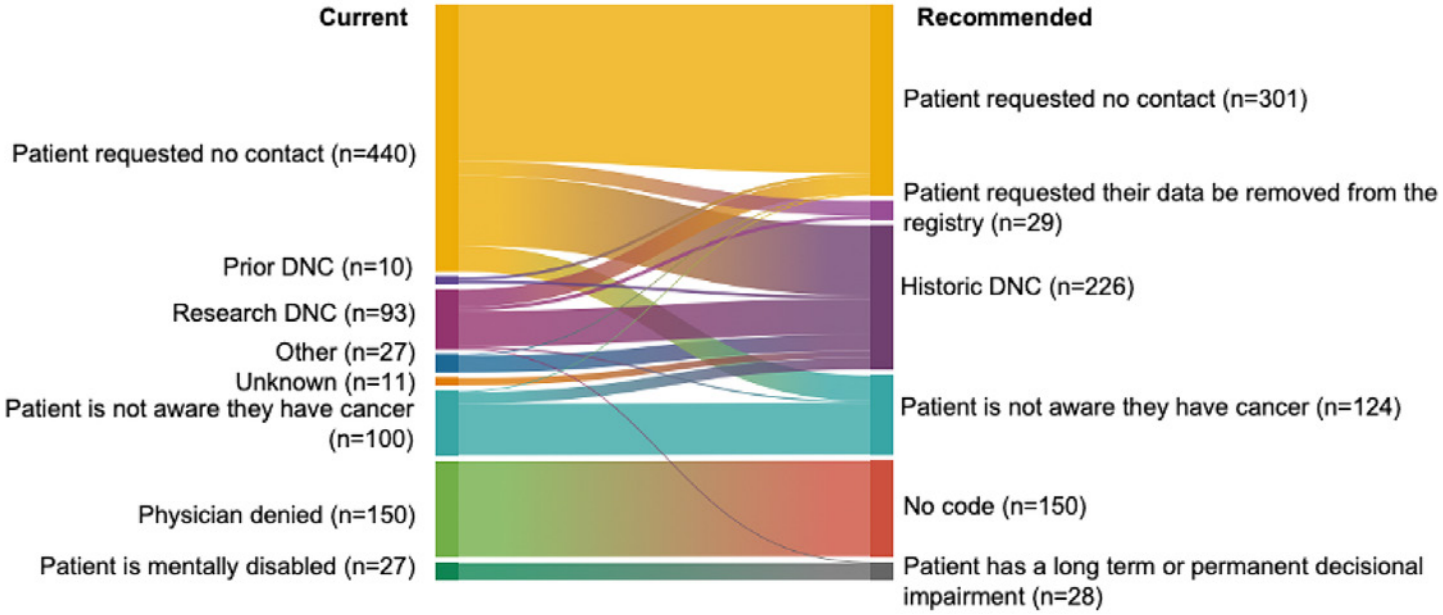
For 29 of the total 858 cases reviewed (3%), we found evidence that the patient wanted their records removed from the registry; 23 were currently coded as *Patient requested no contact*, and 6 were identified from research databases.

Recommendations for DNC Codes and Research Handling in the Utah Cancer Registry

Common reasons for discordance included lack of evidence to support the current DNC flag and reason code, changes in code definitions over time (ie, historic changes), and ambiguity in a physician contact form. These findings informed recommendations for updates to the DNC reason codes to resolve discordance and ensure more appropriate inclusion and exclusion of cancer cases in research.

Recommended changes to the Utah Cancer Registry DNC coding include the addition of 2 new reason codes *(Patient requested their data be removed from registry* and *Historic DNC)* and the retiring of the *Physician denied* code. The *Patient requested their data be removed from the registry* code would be applied to the 29 cases (6 of which were from research databases) with supporting evidence found during the manual review (Figure 1). While cancer is a reportable disease and surveillance is mandated (meaning that records cannot be removed from the Utah Cancer Registry), from a research perspective, we sought to distinguish these cases to exclude their records from all research database linkages as well as research contacts. The *Historic DNC* code would be applied to the 226 cases with a current or prior DNC code in SEER*DMS or a research database, but no evidence found to support any DNC code during the manual review (Figure 1). Our recommendation is that these cases would be included in research linkages but excluded from research contacts. For all 150 cases currently coded as *Physician denied,* we recommend removing the DNC flag and including these records in research linkages and research contacts. Additionally, we will contemporize the code definition for *Patient is mentally disabled* to *Patient has a long-term or permanent decisional impairment.* For cases included because of a prior DNC flag in SEER*DMS or a current DNC flag in a research database (but with no reason code), we recommend the addition of a SEER*DMS DNC flag and reason code based on the evidence found. This recommendation included adding the *Patient requested no contact* code to 5 prior DNC cases and 27 research DNC cases; adding the *Patient is not aware they have cancer* for 3 research DNC cases; and adding the *Patient has a long-term or permanent decisional impairment* for 1 research DNC case. For cases with a current SEER*DMS DNC flag and reason code, if we found no evidence supporting the current DNC reason code, but did find evidence supporting a different code, we recommend updating to the code with supporting evidence. For cases with evidence found supporting the current reason code and also a different code, we recommended

Figure 1. Change in the Distribution of Cases and DNC Reason Codes Between the Current Classification and the Recommended Classification Following Manual Review



DMS, Data Management System; DNC, do not contact.

Cases included because of a prior DNC flag in SEER*DMS or a DNC flag in a research database do not have a current DNC reason code and are included in the "Prior DNC" and "Research DNC" categories, respectively. The alluvial flow from left to right shows the change in distribution of the cases and DNC reason codes based on the evidence found during manual review.

using the code with the highest contact and research restriction. The priority for code recommendation in order of decreasing priority was as follows: (1) *Patient requested their data be removed from registry,* (2) *Patient requested no contact,* and (3) *Patient is not aware they have cancer*. The change in distribution of DNC reason codes following these recommended changes can be found in Figure 2. We propose to use the comprehensive REDCap form that was developed for manual review for future DNC incidents to allow for more informative tracking of future cases.

New NAACCR Standard No Patient Contact Flag

Starting with diagnosis year 2023, the North American Association of Central Cancer Registries (NAACCR) will introduce a new standard data item, *No Patient Contact Flag,*[13] which registries can use to identify cases in which the patient should not be contacted for research purposes. The Utah Cancer Registry will set this new field to, "1 – Patient may NOT be contacted" if a case has any Utah Cancer Registry–defined DNC reason code. In addition to these standardized fields, registries will continue to have autonomy over the use and interpretation of DNC flags and reason codes in SEER*DMS to maintain the reason for DNC.

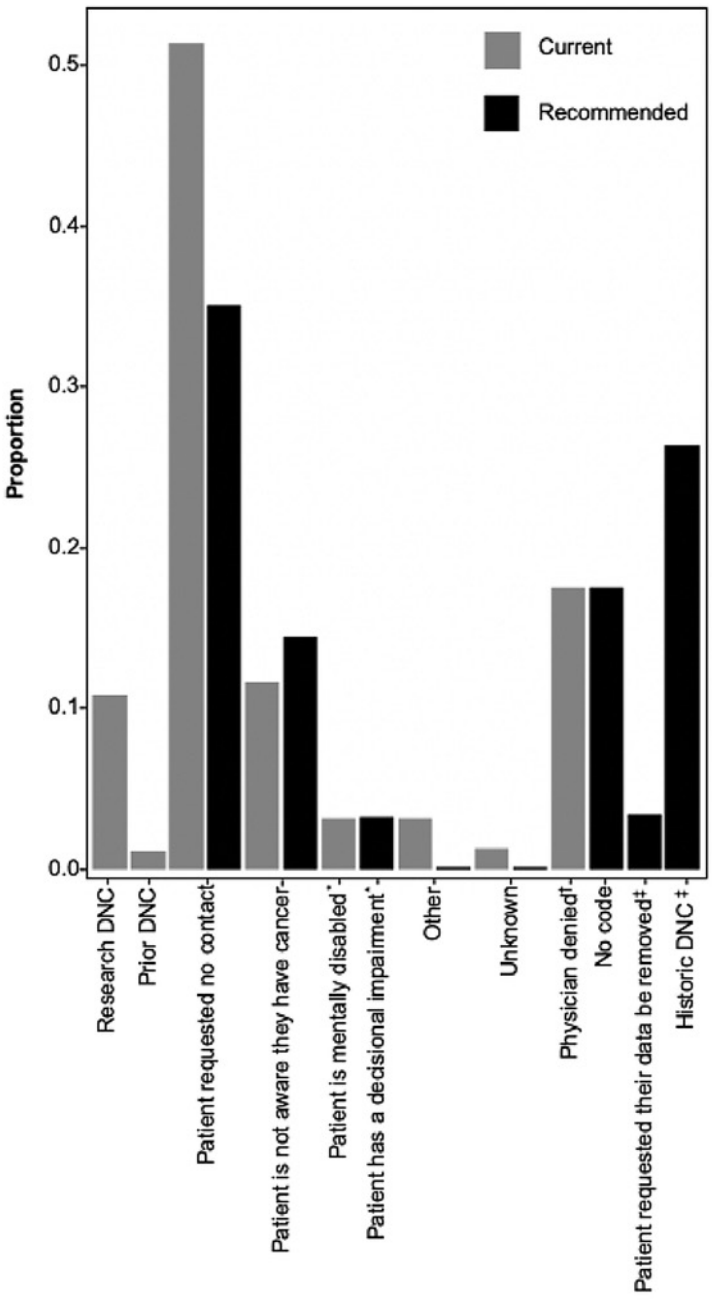
Discussion

Cancer registries have become increasingly responsible for managing appropriate research contacts and record releases. Through a manual review of case records in SEER*DMS, we identified opportunities to improve DNC flag and reason code assignments and research handling in the Utah Cancer Registry. These findings informed the addition of 2 new reason codes, the retiring of an older reason code, and the contemporizing of code definitions. Results from our study suggest that changes in historical practices and code definitions over time may erroneously include or exclude cases from research contacts and research linkages.

While the Utah Cancer Registry now performs passive follow-up through multiple database linkages, including the National Death Index, we historically performed active follow-up by contacting patients and their providers by letter. Notably, the physician letter used during follow-up asked physicians for permission to "locate" rather than "contact" the patient. During this period of active follow-up, the DNC flag was used to indicate patients who either responded to the letter and asked not to be contacted or providers who marked the "do not locate" box. Consequently, discordant cases were identified during manual review that were currently coded as "physician denied" but had no other evidence to support any DNC code besides the checked box on the physician form. The DNC flag and reason codes are now only used for research and are no longer used for any operational processes. Additionally, while other registries require physician permission prior to research contacts, the Utah Cancer Registry discontinued this practice more than 10 years ago. Thus, we recommend retiring the *Physician denied* code.

In addition to changes in the use of DNC codes, the Utah Cancer Registry's practices for recording DNC codes, including incident forms and data management systems, have changed over time. During the manual review, we found a subset of discordant cases currently coded as

Figure 2. Changes in the Proportions of Cases Assigned to DNC Reason Codes Following Manual Review



DMS, Data Management System; DNC, do not contact. The proportions of cases classified by current DNC reason codes in SEER*DMS are shown in grey, and the proportions classified by recommended DNC codes are shown in black (n = 858). Cases included because of a prior DNC flag in SEER*DMS or a DNC flag in a research database do not have a current DNC reason code and are included in the "Prior DNC" and "Research DNC" categories, respectively. All individuals in these two categories were re-classified into recommended codes.

*Recommended updated definition DNC codes: *Patient has a decisional impairment*.

†Recommended retired DNC codes: *Physician denied* (all of the cases in this category were changed to having no code).

DNC that had no evidence to support any DNC code. Many of these discordant cases likely resulted from a data transfer between data management systems that accidentally removed supporting documentation for any DNC reason code. We recommended recoding these cases with a new DNC reason code, *Historic DNC*. We also found cases currently coded as DNC in research databases missing a DNC flag and reason code in SEER*DMS. This finding confirmed the suspicion that some DNC codes may have been missed during the manual transfer from research databases to SEER*DMS.

The manual review allowed us to identify reasons for discordance that would have been missed without a review of scanned images, comment boxes, and research databases. The manual review also allowed us to distinguish patients who wanted their records removed from those who did not want to be contacted, and resolve all cases currently coded as *Unknown* reason. While cancer is a reportable disease and these records will be retained by Utah Cancer Registry for surveillance data sets, distinguishing cases with patients who want their records removed is important for excluding their records from research linkages. These findings informed the addition of a *Historic DNC* code, and a *Patient requested their data be removed from the registry* code. Cases with the *Historic DNC* code will be included in research database linkages but excluded from research contacts, while cases with the *Patient requested their data be removed from the registry* code will be excluded both from research database linkages and research contacts. We will discuss our recommendations for updated research handling of each DNC reason code with a panel of experts in the ethical conduct of human subjects research, including representatives from our institutional review board, before proceeding with any changes.

Limitations of our manual review include that, for the reason code *Patient is not aware they have cancer,* we did not distinguish evidence for the tumor- versus patient-level code. A tumor-level code would exclude the patient from research contacts for that diagnosis only while a patient-level code would exclude the patient from any research contacts. While it would have been possible to distinguish whether the evidence supported the tumor- or patient-level code for some cases, for other cases, this code resulted from a research contact, and it was unclear from the documentation which diagnosis was being referenced. No changes were made to the *Patient is not aware they have cancer* code, and all cases coded at the patient level will continue to be excluded from research contacts. Updated research patient contact procedures define how to set this reason code at the tumor or patient level going forward.

With increasing concerns about data use and privacy,[14] it is important to critically review existing processes for documenting research contacts and record linkages, particularly when potential study participants request not to be contacted. Our manual review of 858 case records in the Utah Cancer Registry took approximately 150 hours (an average of 10 minutes per record) and 25 hours were required for training and meetings with the Utah Cancer Registry managers. Findings from our manual review

highlight the need to contemporize codes and coding practices and the value of manual review to inform this process. A manual review of the DNC flags and reason codes in the Utah Cancer Registry informed updates to the Utah Cancer Registry SEER*DMS reason code definitions and practices that will improve the precision of DNC codes in the Utah Cancer Registry. These policy and procedural updates will ensure that our practices are true to the intended code meaning and respect patient wishes while balancing the need for accurate and complete research data.

References

1. White MC, Babcock F, Hayes NS, et al. The history and use of cancer registry data by public health cancer control programs in the United States. *Cancer*. 2017;123(suppl 24):4969-4976. doi:10.1002/cncr.30905
2. Overview of the SEER program. National Cancer Institute's Surveillance, Epidemiology, and End Results Program website. Accessed September 1, 2022. **https://seer.cancer.gov/about/overview.html**
3. Tucker TC, Durbin EB, McDowell JK, Huang B. Unlocking the potential of population-based cancer registries. *Cancer*. 2019;125(21):3729-3737. doi:10.1002/cncr.32355
4. Gallicchio L, Elena JW, Fagan S, et al. Utilizing SEER cancer registries for population-based cancer survivor epidemiologic studies: a feasibility study. *Cancer Epidemiol Biomarkers Prev*. 2020;29(9):1699-1709. doi:10.1158/1055-9965.EPI-20-0153
5. NPCR standards. Centers for Disease Control and Prevention website. Published August 29, 2022. Accessed September 14, 2022. **https://www.cdc.gov/cancer/npcr/standards.htm**
6. *SEER 2022 Submission Requirements and Guidelines*. The Surveillance, Epidemiology, and End Results (SEER) Program; 2022. **https://seer.cancer.gov/tools/seer.nov22.instructions.pdf**
7. Osborne JD, Wyatt M, Westfall AO, Willig J, Bethard S, Gordon G. Efficient identification of nationally mandated reportable cancer cases using natural language processing and machine learning. *J Am Med Inform Assoc*. 2016;23(6):1077-1084. doi:10.1093/jamia/ocw006
8. Xie F, Lee J, Munoz-Plaza CE, Hahn EE, Chen W. Application of text information extraction system for real-time cancer case identification in an integrated healthcare organization. *J Pathol Inform*. 2017;8:48. doi:10.4103/jpi.jpi_55_17
9. Cancer Reporting Rule, R384-100. Utah Office of Administrative Rules; 2010. **https://uofuhealth.utah.edu/documents/cancer-reporting-rule-july-2021**
10. Beskow LM, Sandler RS, Weinberger M. Research recruitment through US central cancer registries: balancing privacy and scientific issues. *Am J Public Health*. 2006;96(11):1920-1926. doi:10.2105/AJPH.2004.061556
11. Thornton ML, ed. *Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Record Layout Version 18*. North American Association of Central Cancer Registries; 2018.
12. Millar M, Herget K, Sweeney C. *Cancer in Utah: Incidence and Mortality Statistics through 2017*. The Utah Cancer Registry; 2020. **https://uofuhealth.utah.edu/documents/cancer-utah-2017**
13. Item 1854, No Patient Contact Flag. In: Thornton ML, ed. *Standards for Cancer Registries Volume II: Data Standards and Data Dictionary*. Version 23, 24th ed. North American Association of Central Cancer Registries; 2022.
14. McGraw D, Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. *NPJ Digit Med*. 2021;4(1):1-11. doi:10.1038/s41746-020-00362-8

Original Article

Integration of Cancer Screening Data into Routine Cancer Surveillance Systems: A Florida Pilot Project

Monique N. Hernandez, PhD [a]; Gary Levin, BA, CTR [a]; Tara Hylton, MPH [b]; Paran Pordell, DrPH [c]; Arica White, PhD [c]; Virginia Senkomago, PhD [c]

Abstract: Introduction: Integration of screening data into routine cancer surveillance systems can create more robust data systems to inform cancer prevention and control activities. Currently, state central cancer registries do not routinely collect breast and cervical cancer screening data as part of state cancer surveillance activities. Florida conducted a pilot study involving: (1) linkage of breast and cervical cancer screening data from the Florida Breast and Cervical Cancer Early Detection Program (FBCCEDP) from 2009 to 2021 to the Florida Cancer Data System (FCDS) database to capture screening data for matched cancer cases in the FCDS; and (2) evaluation of the feasibility of developing a population-based breast and cervical cancer screening surveillance system by capturing electronic screening data from private health care providers. Methods: In 2018, the FCDS worked with the Florida Department of Health to identify data partners for the 5-year cancer screening pilot project funded by the Centers for Disease Control and Prevention. Engagement of project partners required extensive review of available screening data; data standards and formatting; data transmission schedules and methods; and processing procedures. The FCDS developed a database to integrate multiple source data sets into a single database whereby linkage to the central cancer registry could be performed. Results: The FCDS worked with Suncoast Health Systems, a clinical practice in the Hillsborough region of Florida, and the FBCCEDP to evaluate data availability, standardization of data sets, and data submission schedules for the pilot project. Extensive meetings and data reviews were conducted with both partners in the first phase of the project. The FCDS developed automated data processing procedures to integrate the data into a single cancer screening database and then linked records to the central cancer registry data set. Discussion: Registry collaboration with the FBCCEDP and Suncoast team on data quality and standardization has produced positive results. The project required extensive review of data and produced many lessons learned for development of a cancer screening surveillance system. Our pilot project depended on partnership building, commitment to data quality, and consistency in data submission practices.

Key words: *data standardization, data quality, linkage, Florida*

Introduction

The Florida Cancer Data System (FCDS) is a central cancer registry supported by the National Program of Cancer Registries (NPCR) administered by the Centers for Disease Control and Prevention (CDC). The FCDS was legislatively established by Florida Statute in 1978 and began collecting cancer reports statewide in 1981, becoming a member of the NPCR in 1994.[1,2,3,4] Currently, the FCDS contains approximately 4.6 million unique cancers representing reporting from hospitals, ambulatory surgical centers, radiation treatment facilities, private physician practices, and pathology offices. Linkages to the Florida Bureau of Vital Statistics and hospital discharge records help to improve casefinding, data completeness, and quality. The resulting data set provides hospitals, researchers, public health professionals, and policy makers access to a trove of cancer incidence data. However, lacking from the robust FCDS database are applicable screening data for individuals who have been diagnosed with screen-detectable cancers.

Together, breast and cervical cancer constitute a leading cause of morbidity and death among women in the United States.[5] In 2019, breast cancer was the most common cancer reported and the second most common cause of cancer death among women in Florida, contributing 19,062 (126.54 per 100,000) new cancer cases and more than 3,084 deaths (19.05 per 100,000).[6,7] Despite documented declines, Florida's cervical cancer and mortality rates remain among the highest in the United States. In 2019, there were 1,082 (9.17 per 100,000) new cases of cervical cancer and 354 (2.6 per 100,000) attributable deaths.[5-7]

Early diagnosis of breast cancer through mammography, and cervical cancer through Papanicolaou (Pap) or human papillomavirus (HPV) testing, can prevent disease progression to advanced or invasive stages. Screening is also associated with improved treatment outcomes, higher survival rates, and reduced mortality rates.[8] The United States Preventive Services Task Force (USPSTF) recommends mammography screening for women aged 50 to 74 years every 2 years.[9] The USPSTF also recommends screening for cervical cancer in women aged 21 to 65 years with a Pap test every 3 years or, for women aged 30 to 65 years, with a combination of Pap and HPV tests, or HPV

[a]Florida Cancer Data System, University of Miami Miller School of Medicine. [b]Florida Department of Health. [c]Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Cancer Prevention and Control.

Address correspondence to M. N. Hernandez, PhD. Email: MHernandez5@med.miami.edu.

This publication was supported by cooperative agreement # NU58DP006350 from The Centers for Disease Control and Prevention. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

test alone, every 5 years. Provisions in the Affordable Care Act ensure that nongrandfathered private insurance plans, as well as Medicare and Medicaid expansion states, must cover certain preventive services, including breast and cervical cancer screening, with no cost sharing.[10] For women who are low income, uninsured, or underinsured, the Florida Breast and Cervical Cancer Early Detection Program (FBCCEDP) provides breast and cervical cancer screening, follow-up diagnostic services for abnormal results, and treatment-eligibility referral. In 2020, it was estimated that 79.2% of Florida women aged 50 to 74 years received a mammogram in the past 2 years, while 77.1% of those 21 to 65 years had received a Pap test in the previous 3 years.[11] These rates are comparable to 2020 national breast (mammography) and cervical cancer screening (Pap test) rates (78.3% and 77.9%, respectively).[12,13]

Recognizing that surveillance data systems constitute the basis for the planning, implementation, and evaluation of public health programs, integration of health services data such as screening data into routine cancer surveillance systems can complement incidence, mortality, and survival data.[14] Currently, data collected from women during breast and cervical screenings are not part of the FCDS. Linkage of screening data, for both positive and negative screening results, with the central cancer registry can allow for timely collection of additional information, including who is getting screened, age at screening, number of screenings, screening results, and diagnostic follow-up. Specifically, racial and ethnic information from the screening procedures can help supplement incomplete race data from the registry cancer records. These more comprehensive registry data can be used to inform cancer prevention and control activities such as highlighting common geographical or sociodemographic characteristics of patients who were not screened prior to diagnosis, informing future state cancer screening efforts.

To determine if this comprehensive registry is feasible, we evaluated reporting mechanisms for data availability, data quality, and efficiency in capturing screening data as part of a 5-year award from the CDC. To do this, Florida conducted a pilot study involving: (1) linkage of all breast and cervical cancer screenings from the FBCCEDP for the time period 2009 to 2021 to the FCDS to capture medically-confirmed screening data and self-reported medical history for current cancer patients within the FCDS database; and (2) evaluation of the feasibility of developing a population-based breast and cervical cancer screening surveillance system in Florida by electronically capturing screening data from private health care providers through their electronic health record (EHR) systems.

Methods

Overview of Approach

Florida has adopted a multiphase approach to evaluate and identify an effective mechanism for reporting breast and cervical cancer screening data by (1) Collecting statewide FBCCEP data and regional data from Suncoast Community Health Centers (Suncoast), and (2) linking the consolidated screening database to the FCDS.

Florida Breast and Cervical Cancer Early Detection Program

The FBCCEDP provides breast and cervical cancer screening to women aged 50 to 64 years who are at or below 200% of the federal poverty level and who are underinsured or uninsured. According to the US Census Bureau's Small Area Health Insurance Estimates, approximately 647,000 Florida women between the ages of 50 and 64 years were at or below 200% of the federal poverty level in 2017, with about 25.1% being uninsured.[15] Additionally, women below age 50 years who are symptomatic or have a family history of breast cancer and who meet other eligibility requirements for income and insurance may be screened by the FBCCEDP.[16-18] The FBCCEDP provided over 28,000 screening and diagnostic services to eligible women in fiscal year 2017.[19] The FBCCEDP maintains 16 regional sites serving eligible women from all 67 counties in Florida.

Suncoast Community Health Centers

With support from the FBCCEDP Hillsborough County coordinator, the FCDS established a partnership with Suncoast Community Health Center, a mammography and gynecology center that is a not-for-profit, federally qualified community health center. Accredited by the Accreditation Association for Ambulatory Health Care, Suncoast has served Hillsborough County since 1977. The center provides primary health services to poor, uninsured rural persons and migrant, seasonal agricultural workers in Eastern and Southern Hillsborough County and Lakeland, Polk County. With a network of 10 practices, 3 mobile units, 44 medical providers, and 32 dental providers, Suncoast provided services to 63,250 patients in 2018. The patient population of Suncoast derives from an ethnically diverse population. Further, as a federally qualified community health center, the patient population has disproportionately higher rates of poverty. Approximately 86% of these patients had income at or below 200% of the federal poverty level, with 35% uninsured, 72% from racial and ethnic minority groups, and 12% being agricultural workers.[20] Hillsborough County was selected as the initial pilot site given its large size and racial–ethnic and socioeconomic diversity. Hillsborough County is 28% Hispanic, 15.5% non-Hispanic Black, 50% non-Hispanic White, and 4% Asian. Approximately 17% of people in the county are foreign-born, with a majority from Latin America and Asia (67% and 20%, respectively).[15,21] Hillsborough county constitutes a major urban center and captures a significant population in need of cancer screening services.

Data Submission

Florida Breast and Cervical Cancer Early Detection Program. The FBCCEDP collects the personal health history as well as demographic, screening, testing, diagnosis, and treatment data for women who receive breast and cervical cancer screening services. After a series of discussions in biweekly meetings and review of the FBCCEDP patient reporting form, the FBCCEDP agreed to submit many of the screening and diagnostic follow-up data items captured on the patient reporting form along with key demographic

data required to facilitate linkage to the registry. As part of the pilot project, the FBCCEDP submitted screening data to the FCDS for services and procedures beginning from the year 2009, which was the earliest year that the program implemented an electronic capture data system, through the year 2021.

Following data review, including quality checks, standardization of data values, and deduplication, the data set was consolidated to unique patient-level records. The range of procedures included both screening and diagnostic procedures as well as consultations. The FCDS worked with the FBCCEDP program to identify the procedure types that are specific to screening procedures versus diagnostic procedures to further define the consolidated screening database. These included mammograms, clinical breast examinations (CBEs), Pap tests, and HPV tests. While CDC did not require or recommend the collection of CBEs as a cancer-screening variable for this feasibility study, we included it as a state-specific item to expand the potential for linkage to the central cancer registry.

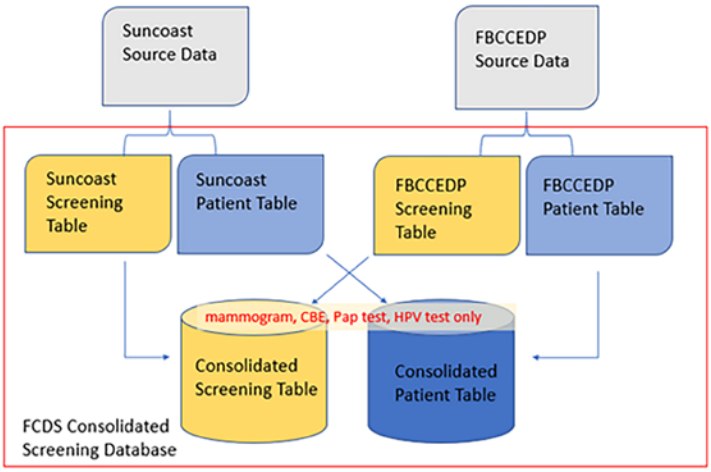
Suncoast. It was determined that the best way to extract breast and cervical cancer screening data was through Suncoast's EHR system. Screening data cover Pap tests, HPV tests, CBEs, and mammograms. Quarterly submissions from Suncoast began in January 2019 and continued through the first half of 2022. Each submission reflected the most recent screening dates by patient. For example, a Pap test performed on January 1, 2019, will remain on each report until the next Pap is performed, when it will be replaced with the latest date on the report for each type of screening.

As FCDS received submissions, the data were parsed by procedure type (CBE, HPV, Pap, mammogram) and results from the screening. The FCDS data import module allowed screening results to be updated as needed if new information was reported.

To date, Suncoast has submitted screening data for women who received screening services between January 2019 and January 2022. Some women who received a single screening during this period also had historical screening data for other screenings included in the data set that went as far back as 2009, but most procedures occurred between 2019 and 2022, as those were the years for which the data were extracted.

Consolidated Screening Surveillance Database. The FCDS team began the design of the consolidated database in November 2020. The first step required developing the related procedures to load and update data from Suncoast and FBCCEDP into respective patient and screening tables (Figure 1). These source-specific tables have their own update logic given structural differences between reporting sources. The second step required consolidation logic to incorporate both sources into a consolidated patient and screening table, with text-to-code conversions where applicable for fields such as race, ethnicity, smoking status, and sex. Where possible, we applied standardized values from the North American Association of Central Cancer Registries (NAACCR).

Figure 1. Consolidated Patient and Screening Database Flowchart



CBE, clinical breast examinations; FBCCEDP, Florida Breast and Cervical Cancer Early Detection Program; FCDS, Florida Cancer Data System; HPV, human papillomavirus; Pap, Papanicolaou.

Each record contains a field to indicate the reporting source. The Florida project team selected 4 procedure types to include in the final consolidated screening database: mammogram, CBEs, Pap test, and HPV test. They included an insert date and last update field to capture the status of each record. The consolidated patient table contains the matched patient identifier and matched date field where records were linked with a patient in the cancer registry tumor database. Tables 1 and 2 show the consolidated fields for the screening and patient tables.

While the FBCCEDP data included both screening and diagnostic procedures, we only included those that fell within the 4 screening criteria: mammogram, clinical breast exams, Pap test, and HPV test. Procedures from Suncoast fell exclusively into 1 of these 4 screening categories.

Cancer registry linkage. A final consolidated FCDS tumor data set was extracted in year 5 of the pilot and included unique female-only patient-level records regardless of diagnosis year; we used the entire FCDS data because some women may have been captured in the registry for any cancer diagnosed prior to or after utilizing the screening services.

The FCDS implemented a probabilistic linkage of the 2 data sets using the R package *fastLink* with a matching threshold of at least 0.98. The linkage variables consisted of social security number, zip code, date of birth, street number, and phonetic encodings of first and last names. The data set of linked cancer patients resulting from the linkage was subsequently used to extract tumor records for each patient. We analyzed the linked data set to generate frequencies of tumor types by site.

The unique patients from the consolidated patient screening database were linked to the consolidated FCDS patient-level database. The tumor records were further limited to diagnosis years 1981-2019, representing complete incidence cases in the registry. No manual review was performed on the linkage. Furthermore, the *fastLink* program includes a deduplication procedure.

Table 1. Consolidated Patient Table Variables
C2P_PATIENT_ID
C2P_SOURCE
C2P_NAME_LAST
C2P_NAME_MIDDLE
C2P_NAME_FIRST
C2P_DOB
C2P_SSN
C2P_SEX
C2P_ADDRESS1
C2P_ADDRESS2
C2P_CITY
C2P_ZIP
C2P_STATE
C2P_BMI
C2P_HEIGHT
C2P_WEIGHT
C2P_HEIGHT_TEXT
C2P_WEIGHT_TEXT
C2P_VITAL_STS_DATE
C2P_SMOKING_STATUS
C2P_TOBACCO_STS_DATE
C2P_INSURANCE
C2P_RACE
C2P_ETHNICITY
C2P_INSERT_DATE
C2P_LAST_UPDATED
C2P_COUNTY
C2P_MATCHED_PATIENT_ID
C2P_MATCHED_DATE

Table 2. Consolidated Screening Table Variables
C2S_PATIENT_ID
C2S_SOURCE
C2S_SCREENING_TYPE
C2S_SCREENING_DATE
C2S_SCREENING_RESULT
C2S_INSERT_DATE
C2S_LAST_UPDATE

Results

FBCCEDP

Upon consolidation of the FBCCEDP source data to include only those screenings that fell within the 4 main screening categories, the total number of unique patients totaled 114,713 patients and 331,414 screenings (49% mammogram, 31% CBE, 16% Pap test, and 4% HPV test). Approximately 92% of all patients were between 50 and 64 years at the time of the procedure, 39% were Non-Hispanic White, 22% were Non-Hispanic Black, and 34% were Hispanic.

Suncoast

The Suncoast data set represented a total of 19,308 unique female patients and 53,553 screening procedures. Pap tests comprised most of cancer screening procedures (40%), followed by HPV tests (29%), mammograms (18%), and CBE (13%). Of the 19,308 unique patients, 63% were Hispanic, 26% Non-Hispanic White, and 9% Non-Hispanic Black. Many patients were aged 24-49 years (65%), and 35% of individuals represented the 50 years or older age group. The data extraction process from Suncoast pulls records based on the current years' screenings, but also includes the date of last screening across screening modalities. Therefore, screenings have been captured that predate the year of data extract.

Consolidated Screening Database

The design and functioning of the consolidated screening database have been effectively automated and handles the unique formats of each data source to produce a standardized analytic data set. There are 2 main analytic tables; patient-level information is consolidated by source into a single patient table, and screening procedures are also consolidated by source (Suncoast and FBCCEDP). The final patient and screening tables combine sources into a single analytic set. Tables 3 and 4 describe the total number of consolidated and unique patient records and screenings combined by source and demographic characteristics.

Cancer Registry Linkage Results

The results of the tumor linkage for Florida malignant breast and cervical cancers resulted in a total of 5,602 linked patients and 6,006 patient tumors (there can be multiple tumors per patient) diagnosed between 1981 and 2019. The linkage included a de-duplication component, did not use blocking, and did not include manual review. Given that there were known duplicate patients in the consolidated screening patient table, as described previously, the de-duplication was helpful in eliminating many of these instances. The final linked tumor data set excluded out of state diagnosis, and removed tumors diagnosed prior to 1981 and after 2019. The cancer registry was established in 1981 and the most recently published data were for the 2019 diagnosis year. This ensures higher data quality and validation. Therefore, we included a total of 6,006 breast and cervical tumors diagnosed among 5,602 patients in the final analytic data set. Most linked tumors consisted of cancers of the breast (93%) with fewer cervical cancers (7%)

Table 3. Consolidated and Unique Patient Records, Suncoast (2019–2022) and FBCCEDP (2009–2021)

	Combined		*Suncoast*		*FBCCEDP*	
	n	**%**	**n**	**%**	**n**	**%**
Total patient records	134,021	100	19,308	100	114,713	100
Race/ethnicity						
Hispanic	52,037	39	12,161	63	39,876	35
Non-Hispanic White	49,835	37	4,962	26	44,873	39
Non-Hispanic Black	27,577	21	1,780	9	25,797	22
Non-Hispanic AI/AN/Other*	2,784	2	13	<1	2,771	2
Unknown	1,788	1	392	2	1,396	1
Smoking Status						
Never smoker	98,724	74	14,758	76	83,966	73
Current	25,601	19	2,118	11	23,483	20
Former smoker	2,432	2	2,432	13	0	0
Unknown	7,264	5	0	0	7,264	6

AI/AN, American Indian/Alaska Native; FBCCEDP, Florida Breast and Cervical Cancer Early Detection Program.

Table 4. Consolidated Breast and Cervical Cancer Screening Procedures, Suncoast (2019–2022) and FBCCEDP (2009–2021)

	Combined		*Suncoast*		*FBCCEDP*	
	n	**%**	**n**	**%**	**n**	**%**
Total screening records	384,967	100	53,553	100	331,414	100
Procedure						
Mammogram	173,888	45	9,493	18	164,395	50
Clinical breast examination	108,597	28	7,293	13	101,304	30
Pap test	72,822	19	21,365	40	51,457	16
HPV	29,660	8	15402	29	14,258	4
Age at time of screening (y)						
<20	53	<1	0	0	53	<1
20–29	7,230	2	6,162	12	1,068	<1
30–39	20,347	5	16,152	30	4,195	1
40–49	25,923	7	14,504	27	11,419	3
50–59	230,524	59	11,592	22	218,932	66
60–69	99,723	26	4,863	9	94,860	29
>70	1,167	<1	280	<1	887	<1

FBCCEDP, Florida Breast and Cervical Cancer Early Detection Program; Pap, Papanicolaou.

(Table 5). It is worth noting that the source of screening data skew toward individuals of lower socio-economic status. Therefore, the distribution of covariates is not representative of all diagnosed breast and cervical cancers in the FCDS.

Limitations

There were multiple challenges throughout the pilot screening project that are worth noting. First, establishing data sharing partnerships requires a considerable time commitment and investment in effort to meet with health agencies, private practices, clinicians, and public health officers. While connections to internal and partnered agencies within the Department of Health were more easily facilitated, relationships with private practices and clinicians required more effort given less familiarity with the central cancer registry's legislative mandates. For example, a prospective clinical practice declined to work with the pilot project citing concerns around the releasing of patient identifiable information (PII) and potentially violating patient confidentiality. Although the current state cancer registry legislation and a data use agreement with the health care system allowed the project team to obtain cancer screening

Table 5. Linked Tumors by Site with Consolidated Unique Patient Table (n = 6,006), FCDS, 1981–2019						
Site	*Combined*		*Suncoast*		*FBCCEDP*	
	n	%	n	%	n	%
Breast	5,595	93	113	84	5,482	93
Cervix uteri	411	7	21	16	390	7
Total	6,006	100	134	100	5,872	100

FBCCEDP, Florida Breast and Cervical Cancer Early Detection Program; FCDS, Florida Cancer Data System.

data for this feasibility project, there were concerns about whether the legislation covered cancer screenings not linked to a diagnosis or treatment of cancer. Developing a statewide cancer screening surveillance system may require policy revisions, including more specific legislative language that requires reporting of all cancer screening data, with or without a cancer diagnosis and prospectively, to circumvent any barriers to acquiring data in the future.

Secondly, there are no data standards for capturing screening data from the EHR. Each provider and program use different file layouts. The FBCCEDP data set contains multiple patient records if more than 1 procedure was performed, while the Suncoast data set structure contains patient data on a single record with multiple procedures listed in different fields. A statewide roll-out would require a data dictionary for consistent reporting and assimilation to the standard format by each provider. Providers use different EHR software resulting in the various file layouts and formatted data. While some data elements are coded into distinct categories with defined values, others are entered in as free text or without formatting standards. The business flow and logic of collecting and maintaining the data sets in the EHR are often combined with ad hoc data cleaning and recoding, resulting in less standardized data sets. A statewide roll-out would require a data dictionary defining consistent coding standards by each provider.

Third, EHR system upgrades have contributed to delays in data extraction and submission. Going forward, software vendor changes could also represent a potential barrier to receiving timely data from other providers.

Lastly, the COVID-19 pandemic shifted organizational priorities across the state, which directly and indirectly impacted the pilot project in terms of the availability of providers to dedicate efforts toward data review and submission. While it did not stall progress on the project completely, it presented additional challenges that affected project implementation and efficiencies.

Discussion

The 5-year cancer screening surveillance feasibility pilot project presented Florida with a valuable learning experience and provided an opportunity to create a road map for inclusion of cancer screening surveillance data within the central cancer registry. The linked screening and tumor database can be analyzed to address specific screening to diagnosis research questions that may include identification of delayed screenings, time from screening to diagnosis, screening disparities and associated advanced stage tumors, among other inquiries that inform evidence-based policy and program decision making. The addition of other cancer screening modalities such as low dose computed tomography (CT) scans for lung cancer, and colonoscopies for colorectal cancer can be considered in the expansion of this project.

Prior studies that linked breast cancer screening with cancer registry data were conducted over 20 years ago and focused on acquiring data from a screening and provider or health insurance claims database.[22,23] Although these studies demonstrated the feasibility of acquiring breast cancer screening data, the quality and completeness of variables available for linkage with registry varied. Additionally, prior studies concluded that performing cancer screening and registry data linkage was resource and time intensive and posed data privacy and security issues. A recent study by Heins and colleagues linked national cancer registry data with a small percentage (10%) of patient-level primary care EHRs without reliance on a unique identifier; this limited data quality and completeness and may have resulted in false data linkages.[24]

Integral to the overall completion and success of the feasibility pilot was ensuring that annually proposed planned activities were evaluated, identifying facilitators in accomplishing planned activities and discovering challenges and barriers, which delayed completion of proposed activities and consequently the development of new activities. Important factors for success involve facilitators at the state and county levels to connect the cancer registry with potential data partners. Without these support networks, identifying appropriate data sources and partners would have been extremely difficult. Based on Florida's experience implementing this project, expanding this project to additional registries would require each of the barriers to be fully addressed. A formalized and standardized process for data capture, formatting, and submission, as well as the development of a statewide data dictionary, is vital to planning and implementing a cancer screening surveillance system within an existing central cancer registry infrastructure.

References

1. Florida State Statute 381.0031—Public Health: General provisions; epidemiological research; report of diseases of public health significance to department. 2021.
2. Florida State Statute 395—Public Health: Hospital licensing and regulation. 2021.
3. Florida State Statute 408.07, rule 64D-3—Public Health: Health care administration; control of communicable diseases and conditions which may significantly affect public health. 2021.

4. Florida State Statute 483—Regulation of Professions and Occupations: Health testing services. 2021. **https://www.flsenate.gov/Laws/Statutes/2016/Chapter483/All**
5. US Cancer Statistics Working Group. US Cancer Statistics Data Visualizations Tool, based on November 2017 submission data (1999–2015): US Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Centers for Disease Control and Prevention and National Cancer Institute. 2018;6.
6. Statistics and dashboards. Florida Cancer Data System website. Accessed April 21, 2022. **https://fcds.med.miami.edu/inc/statistics.shtml**
7. Certificates and registries. Florida Department of Health website. Accessed April 21, 2022. **https://www.floridahealth.gov/certificates/**
8. Henley SJ, King JB, German RR, Richardson LC, Plescia M; Centers for Disease Control and Prevention (CDC). Surveillance of screening-detected cancers (colon and rectum, breast, and cervix)—United states, 2004–2006. *MMWR Surveill Summ*. 2010;59(9):1-25.
9. Final recommendation statement: breast cancer: screening. US Preventive Services Task Force website. Published January 11, 2016. Accessed May 7, 2019. **https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/breast-cancer-screening**
10. White A, Thompson TD, White MC, et al. Cancer screening test use—United States, 2015. *MMWR Morb Mortal Wkly Rep*. 2017;66(8):201-206.
11. Florida behavioral risk factor data. Florida Department of Health website. Accessed August 4, 2022. **http://www.flhealthcharts.com/charts/Brfss.aspx**
12. Screening and risk factor table, 2020—mammogram in past 2 years, ages 50–74. Accessed November 28, 2022. State cancer profiles website of the National Cancer Institute and the Centers for Disease Control and Prevention. **https://statecancerprofiles.cancer.gov/risk/index.php?stateFIPS=00&topic=women&risk=v05&race=00&sex=2&type=risk**
13. Screening and risk factor table, 2020—pap smear in past 3 years, no hysterectomy, ages 21–65. State cancer profiles website of the National Cancer Institute and the Centers for Disease Control and Prevention. Accessed November 28, 2022. **https://statecancerprofiles.cancer.gov/risk/index.php?topic=women&risk=v17&race=00&type=risk&sortVariableName=default&sortOrder=default**
14. Ryerson AB, Massetti GM. CDC's public health surveillance of cancer. *Prev Chronic Dis*. 2017;14:E39.
15. Small area health insurance estimates. United States Census Bureau website. Accessed May 7, 2019. **https://www.census.gov/data-tools/demo/sahie/#/**
16. Ryerson AB, Benard VB, Major A. *National Breast and Cervical Cancer Early Detection Program: Early Detection: Summarizing the First 12 Years of Partnerships and Progress Against Breast and Cervical Cancer: 1991–2002 National Report*. Centers for Disease Control and Prevention; 2004.
17. White A, Miller J, Royalty J, et al. Clinical outcomes of mammography in the national breast and cervical cancer early detection program, 2009–2012. *Cancer Causes Control*. 2015;26(5):723-732.
18. Lantz PM, Mullen J. The national breast and cervical cancer early detection program: 25 years of public health service to low-income women. *Cancer Causes Control*. 2015;26(5):653-656.
19. *The Florida Breast Cancer Early Detection and Treatment Referral Program Report: Florida Breast and Cervical Cancer Early Detection Program*. Florida Department of Health; 2019. **https://www.floridahealth.gov/diseases-and-conditions/cancer/breast-cancer/brst-cerv-cancer-legreport-2019-0101.pdf**
20. Health center data. Health Resources & Services Administration website. Accessed May 7, 2019. **https://bphc.hrsa.gov/uds/datacenter.aspx?q=d&bid=041750&state=FL&year=2017**
21. City and town population totals: 2010–2019. United States Census Bureau website. Accessed April 29, 2021. **https://www.census.gov/data/datasets/time-series/demo/popest/2010s-total-cities-and-towns.html**
22. Van Hal G, Thibaut A, Matthyssen M, Weyler J. Linking a breast cancer screening data base with a cancer registry in Antwerp, Belgium. *Arch Public Health*. 2000;58:307-319.
23. Doebbeling BN, Wyant DK, McCoy KD, et al. Linked insurance-tumor registry database for health services research. *Med Care*. 1999;37(11):1105-1115.
24. Heins MJ, de Ligt KM, Verloop J, Siesling S, Korevaar JC. Opportunities and obstacles in linking large health care registries: the primary secondary cancer care registry-breast cancer. *BMC Med Res Methodol*. 2022;22(1):1-7.

Original Article

The Central Brain Tumor Registry of the United States Histopathological Grouping Scheme Provides Clinically Relevant Brain and Other Central Nervous System Categories for Cancer Registry Data

Quinn Ostrom, PhD, MPH[a,b,c]; Carol Kruchko, BA[a]; Corey Neff, MPH[a,b]; Albert Firth, BA[d]; Recinda Sherman, MPH, PhD, CTR[e]

Abstract: Background: Brain and other central nervous system (CNS) tumors are a heterogenous collection of tumors, but they are generally reported in local and national cancer statistics as a single, large category. Although the collection of nonmalignant brain and other CNS tumors has been mandated since diagnosis year 2004, these tumors are often excluded from standard statistical reports on cancer despite their burden on populations in the United States and Canada. The Central Brain Tumor Registry of the United States (CBTRUS) historical and current histopathological grouping schemes have been developed in collaboration with neuropathologists to capture the diversity of these tumors in clinically relevant categories. The goal of this analysis was to test a new recode variable based on the CBTRUS histopathology grouping prior to releasing the variable for use in the North American Association of Central Cancer Registries (NAACCR) Cancer in North American (CiNA) data sets and by individual cancer registries. Methods: The CBTRUS histopathology grouping scheme variable was created and implemented in an evaluation CiNA data set. The accuracy of the variable's categories was evaluated. Counts and incidence rates were calculated using SEER*Stat. Results: Overall, 481,650 cases of brain and other CNS tumors meeting the CBTRUS definition were identified for diagnosis years 2015–2019 in the CiNA data set for the US and Canada, making these the sixth-most-common tumor as a group. Of the brain and other CNS tumor cases, approximately 29% were malignant (behavior code /3 in the *International Classification of Diseases for Oncology,* 3rd edition [ICD-O-3]) while about 71% were nonmalignant (ICD-O-3 behavior code /0 or /1). The overall age-adjusted annual incidence rate (AAAIR) of brain and other CNS tumors was 24.44 per 100,000 (95% CI, 24.37–24.51). The most common histopathologies were meningioma, of which approximately 99% were nonmalignant (AAAIR, 9.09 per 100,000; 95% CI, 9.05–9.13); tumors of the pituitary, of which about 99% were nonmalignant (AAAIR, 4.28 per 100,000; 95% CI, 4.25–4.31); and glioblastoma, of which 100% were malignant behavior (AAAIR, 3.20 per 100,000; 95% CI, 3.18–3.22). Conclusions: Brain and other CNS tumors make up an extremely diverse category that contributes substantially to the cancer burden in North America. The CBTRUS histopathology grouping variable provides clinically relevant groupings for analysis of these tumors in the NAACCR CiNA as well as by individual central cancer registry groups. We encourage the use of this variable to support more detailed analysis of this important group of tumors.

Key words: *brain and central nervous system tumors, Central Brain Tumor Registry of the United States, World Health Organization classification of tumors of the central nervous system*

Introduction

Brain and other central nervous system (CNS) tumors are a heterogenous group of tumors with more than 100 valid *International Classification for Diseases for Oncology,* 3rd edition (ICD-O-3) histopathology codes for these sites. While these histopathologies have extremely disparate incidence and outcomes, most cancer statistics report these cases in bulk based on site/topology codes alone and without any stratification by important histopathologic groups. Additionally, approximately 70% of tumors occurring at CNS sites have benign or borderline behavior and may not be included in overall cancer statistics. For most cancers, malignant behavior is the most important predictor of mortality. But for brain and other CNS tumors, location is a primary predictor of significant morbidity. The importance of including all primary brain tumors in cancer reporting was evidenced by Schoenberg et al in the 1970s.[1] However, most cancer statistics reporting groups (eg, SEER*Explorer, CiNA Explorer) continue to report only malignant brain and other CNS tumors or report incidence of nonmalignant brain and other CNS tumors separately, thereby underestimating the burden of newly diagnosed primary brain and other CNS tumors.

To address this and to provide clinically relevant statistics on brain and other CNS tumors, the Central Brain Tumor Registry of the United States (CBTRUS) was founded in 1992.[2] CBTRUS provided incidence rates from a subset of central cancer registries, resulting in support for

[a]Central Brain Tumor Registry of the United States, Hinsdale, Illinois. [b]Duke University School of Medicine, Durham, North Carolina. [c]Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina. [d]Information Management Systems, Inc, Rockville, Maryland. [e]North American Association of Central Cancer Registries, Inc, Springfield, Illinois.

Address correspondence to Recinda Sherman, NAACCR, 2050 W Iles Ave, Suite A, Springfield, IL 62704-4194. Email: rsherman@naaccr.org.

Funding details for the Central Brain Tumor Registry of the United States can be found at **https://www.cbtrus.org**.

the inclusion of "benign" brain tumors by the Centers for Disease Control and Prevention (CDC)'s National Program of Cancer Registries (NPCR).[3] As a result of the passage of Public Law 107-260 in 2002 (the Benign Brain Tumor Cancer Registries Amendment Act), benign and borderline brain tumors were collected starting January 1, 2004.[4] This law based the collection of brain and other CNS tumors on a site definition rather than a behavior definition as put forth by the currently defunct National Coordinating Council of Cancer Registries and has been agreed upon by members of the cancer surveillance community in concert with representatives from the brain tumor clinical and research communities in Chicago in November 2000 (Consensus Conference 1).[5] Since 2009, NPCR has provided data from central cancer registries on brain and other CNS tumors to the CBTRUS, which CBTRUS combines with central cancer registry data from the National Cancer Institute (NCI)'s Surveillance, Epidemiology, and End Results (SEER) Program for its annual reports and publications. Data from 52 central cancer registries (48 from NPCR and 4 from SEER) presented by behavior, histopathology, sex, age, race, Hispanic ethnicity, and geographic location are included in the 2022 CBTRUS statistical report with data from 2015–2019.[6]

In order to reduce the over 100 histopathologic codes to the clinically appropriate analytic groupings, CBTRUS has worked in collaboration with multiple consulting neuropathologists to develop the CBTRUS histopathologic grouping scheme, which is updated when major changes are made to brain and other CNS tumor classification. In 2016, the *WHO Classification of Tumours of the Central Nervous System*[7] was updated to reflect substantial changes in classification and diagnostic practice that significantly impacted the collection and reporting of CNS tumors. This led to the development of new ICD-O-3 codes, as well as to the development of a new variable (Brain Molecular Markers Site Specific Data Item [SSDI] #3816) by the North American Association of Central Cancer Registries (NAACCR) to fully capture new entities defined in the updated classification.[8] These new codes and SSDIs were implemented beginning in diagnosis year 2018. In addition, these changes led to significant updates to both the SEER and CBTRUS histopathologic grouping schemes for tumors of the CNS.[9,10]

To facilitate broader use of clinically relevant histopathology groupings for tumors of the brain and CNS by the broader cancer registry community, CBTRUS has worked with NAACCR to provide the CBTRUS histopathology recode within NAACCR Cancer in North America (CiNA) data sets, including CiNA Public Use.

There are several notable differences in the definition of brain and other CNS tumors between different reporting groups that should be noted when using site-specific recodes. SEER, NAACCR, and NPCR define brain and other CNS tumors in their reporting as tumors located in the brain, meninges, and other parts of the CNS (ICD-O-3 site codes: C70.0–9, C71.0-9, and C72.0–9), with the exclusion of lymphoma and leukemia histopathologies (ICD-O-3 codes 9590–9989) occurring at those sites. CBTRUS includes the brain, meninges, other CNS tumors, pituitary, craniopharyngeal duct, and pineal gland (ICD-O-3 site codes C70.0–9, C71.0–9, C72.0–9, C73.3–5), as well as olfactory tumors of the nasal cavity (ICD-O-3 site code C30.0, ICD-O-3 histopathology code 9522–9523 only) and lymphomas and leukemias occurring at brain and CNS sites. The inclusion of tumors of the pituitary, craniopharyngeal duct, and pineal gland as well as primary CNS lymphoma as CNS tumors is in line with the World Health Organization (WHO) classification of CNS tumors, but their inclusion in a site-specific recode is unique to the CBTRUS grouping scheme. CBTRUS also includes all primary brain and other CNS tumors irrespective of behavior. Brain tumors with ICD-O-3 behavior codes /0 (benign) and /1 (borderline) are referred to as *nonmalignant brain tumors*. Many reports using the term "brain tumor" or "brain cancer" may be restricted to malignant brain tumors only, despite these tumors representing only about 30% of primary brain tumors.[6]

CBTRUS revised its histopathology groupings to align with the 2016 WHO classification in 2021 while retaining histopathologies that were deemed obsolete for historical comparisons. This process and its potential effect on reporting are detailed at length in Waite, et al.[9] The CBTRUS histopathology recode is described in Table 1. CBTRUS also collaborated with SEER to use a new recode that grouped some of the more common histologies into the histopathology groupings used in the 2016 WHO classification, and this recode is now available in both publicly available SEER and NAACCR data sets.[10] Table 2 provides an overview of classification differences between the SEER and CBTRUS brain and CNS recodes. The objective of this report is to clearly describe the CBTRUS recode and provide guidance for its use by the broader cancer registry community. CBTRUS strongly encourages the use of brain and CNS tumor-specific recodes based on histopathology as opposed to site-based recodes when analyzing brain and CNS data collected by cancer registries for the production of clinically relevant statistics.

Methods

The NAACCR CiNA data set used for this analysis includes data from 67 central cancer registries: 56 from the United States (49 states [excluding Nevada due to data quality issues], the District of Columbia, and Puerto Rico, and 5 regions [California, excluding the Greater Bay area and Los Angeles, plus 4 metropolitan areas]) and 11 from Canada (9 provinces and 2 territories).[11] A special evaluation version of a CiNA Public Use data set[12] was used, which included a recode variable based on a standard SEER*Stat variable, Site Recode ICD-O-3/WHO 2008, and the ICD-O-3 histology codes. The recode variable was based on the Site Recode ICD-O-3/WHO 2008 equal to *Brain and Other Nervous System* (ICD-O-3 codes C70.0-72.9), and histology was aligned with the CBTRUS groupings. For the purposes of this evaluation, the CBTRUS brain and CNS site definition was used (ICD-O-3 site codes C70.0–9, C71.0–9, C72.0–9, C73.3–5, as well as C30.0) for ICD-O-3 histopathology code 9522–9523 only and behavior codes of /0, /1, and /3.

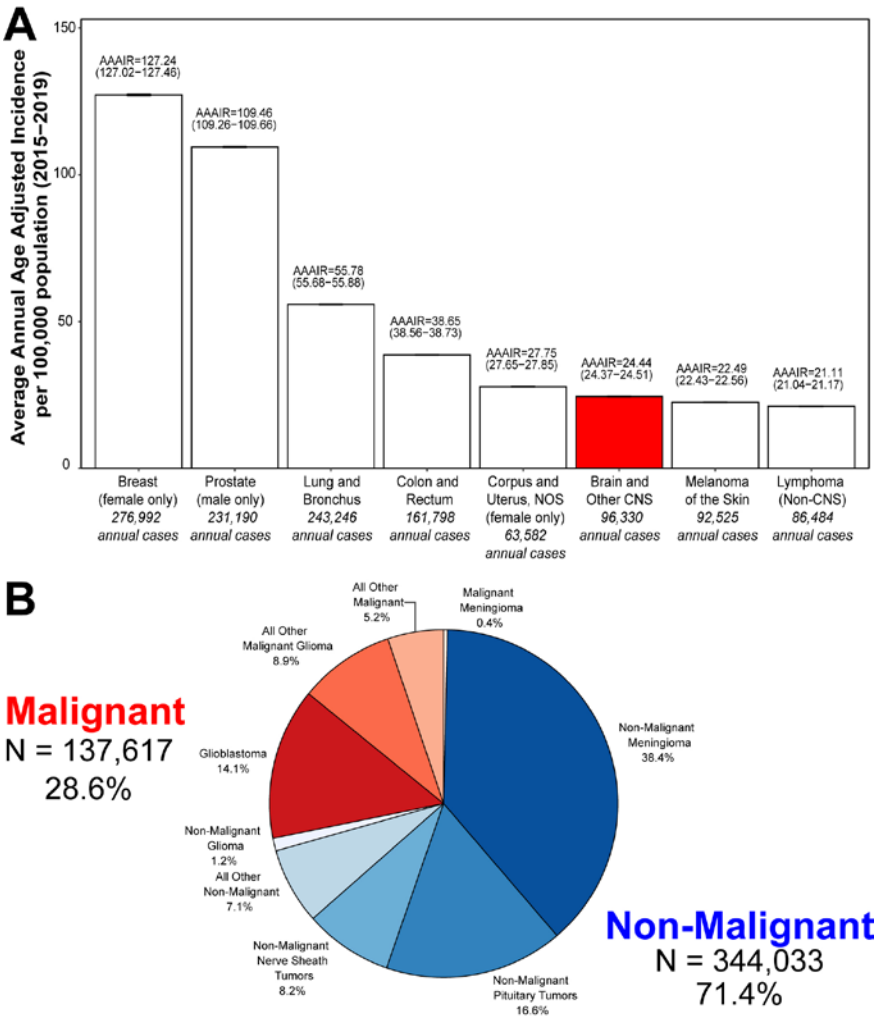
Table 1. Central Brain Tumor Registry of the United States (CBTRUS), 2021 Brain and Other Central Nervous System Tumor Histopathology Groupings (Based on 2016 WHO Classification)

Histopathology	*ICD-O-3[a] Histopathology Codes[b]*
Diffuse astrocytic and oligodendroglial tumors	
Diffuse astrocytoma*	9381, 9400, 9410, 9411, 9420, 9442/1
Anaplastic astrocytoma*	9401
Glioblastoma*	9440, 9441, 9442/3, 9445[c]
Oligodendroglioma*	9450
Anaplastic oligodendroglioma*	9451, 9460
Oligoastrocytic tumors*	9382
Other astrocytic Tumors	
Pilocytic astrocytoma*	9421, 9425[c]
Unique astrocytoma variants*	9384, 9424, 9431[c]
Ependymal tumors*	9383, 9391 (excluding site C75.1 for behavior /1), 9392–9394, 9396[c]
Other gliomas	
Glioma malignant, NOS*	9380, 9385[c]
Other neuroepithelial tumors*	9423, 9430, 9444
Neuronal and mixed neuronal-glial tumors*	8680, 8681, 8690, 8693, 9412, 9413, 9490, 9492 (excluding site C75.1), 9493, 9505, 9506, 9509[c], 9522 (site C30.0 only), 9523 (site C30.0 only)
Choroid plexus tumors	9390
Tumors of the pineal region	9360, 9361, 9362, 9395[c]
Embryonal tumors	8963, 9364, 9470–9478c, 9480, 9500, 9501/3, 9502/3, 9508
Medulloblastoma	9470–9472,9474–9478
Atypical teratoid rhabdoid tumor	9508
Other embryonal tumors	8963, 9364, 9473, 9480, 9500, 9501, 9502
Tumors of cranial and paraspinal nerves	
Nerve sheath tumors	9540, 9541, 9550, 9560, 9561, 9570, 9571
Other tumors of cranial and paraspinal nerves	9562, 9563
Tumors of meninges	
Meningioma	9530–9535, 9537–9539
Mesenchymal tumors	8324, 8710, 8711, 8800–8806, 8810, 8811, 8815, 8821, 8824, 8825, 8830, 8831, 8835, 8836, 8840, 8850–8854, 8857, 8861, 8870, 8880, 8890, 8897, 8900–8902, 8910, 8912, 8920, 8921, 8935, 8990, 9040, 9120, 9125, 9130, 9131, 9133, 9136, 9150, 9161, 9170, 9180, 9210, 9220, 9231, 9240, 9241, 9243, 9260, 9370–9373
Primary melanocytic lesions	8720, 8728, 8770
Other neoplasms related to the meninges	None
Lymphomas and hematopoietic neoplasms	
Lymphoma	9590, 9591, 9596, 9650–9655, 9659, 9661–9665, 9667, 9670, 9671, 9673, 9675, 9680, 9684, 9687, 9688, 9690, 9691, 9695, 9698, 9699, 9701, 9702, 9705, 9712, 9714, 9715, 9719, 9724, 9727–9729, 9735, 9737, 9738, 9750, 9751, 9755, 9756, 9811–9819, 9823, 9826, 9827, 9831, 9832, 9837, 9861, 9866, 9930, 9965, 9966, 9967, 9970, 9971, 9975
Other hematopoietic neoplasms	9731, 9733, 9734, 9740, 9741, 9749, 9752–9754, 9757–9758, 9759, 9760, 9766, 9860,
Germ cell tumors	8440, 9060, 9061, 9064, 9065, 9070–9072, 9080–9083, 9084/3, 9085, 9100, 9101
Tumors of sellar region	

Table 1, cont. Central Brain Tumor Registry of the United States (CBTRUS), 2021 Brain and Other Central Nervous System Tumor Histopathology Groupings (Based on 2016 WHO Classification)	
Histopathology	*ICD-O-3[a] Histopathology Codes[b]*
Tumors of the pituitary	8040 (site C75.1 only), 8140 (site C75.1 only), 8146 (site C75.1 only), 8246, 8260 (site C75.1 only), 8270–8272, 8280, 8281, 8290, 8300, 8310, 8323, 9391/1 (site C75.1 only), 9432[c] (site C75.1 only), 9492 (site C75.1 only), 9580, 9582
Craniopharyngioma	9350–9352
Unclassified tumors	
Hemangioma	9121–9123, 9133, 9140
Neoplasm, unspecified	8000–8005, 8010, 8020, 8021
All other	8320, 8452, 8713, 8896, 8963, 8980, 9084/0, 9173, 9363, 9503

NOS, not otherwise specified; WHO, World Health Organization. [a] *International Classification of Diseases for Oncology*, 3rd edition, 2000. World Health Organization, Geneva, Switzerland. [b] See the CBTRUS website for additional information about the specific histopathology codes included in each group: **https://www.cbtrus.org**. [c] Added starting with diagnosis year 2018.
* All or some of this histopathology is included in the CBTRUS definition of gliomas, including ICD-O-3 histopathology codes 9380–9384 and 9391–9460.

Figure 1. (A) Average Annual Age-Adjusted Incidence Rates (AAAIRs) with 95% CIs of All Primary Brain and Other Central Nervous System Tumors in Comparison to Top 8 Highest Incidence Cancers and (B) Distribution of All Primary Brain and Other Central Nervous System Tumors by Behavior and Most Common Histopathologies, 2015–2019 NAACCR CiNA



CNS, central nervous system; NAACCR CiNA, North American Association of Central Cancer Registries Cancer in North America; NOS, not otherwise specified.
[a] Rates are per 100,000 and are age-adjusted to the 2000 US standard population.
[b] Percentages may not add up to 100% due to rounding.

Counts, average annual age-adjusted incidence rates (AAAIR) per 100,000 population, and 95% CIs by CBTRUS histopathology groupings, behavior, and sex were generated for diagnosis years 2015–2019 for all registries included in the special NAACCR CiNA data set using SEER*Stat 8.4.0.[13]

Results

Overall, 481,650 cases (96,330 cases annually on average) of brain and other CNS tumors meeting the CBTRUS definition were identified in the 2015–2019 NAACCR CiNA data set, including both US and Canadian central cancer registries (Table 3). This is about 25% higher than when the SEER brain and CNS site recode definition is used to identify tumors of the brain and CNS (383,949 total cases for the 5-year period, for an annual average of 76,790 cases), and about 282% higher than when cases are limited to those with malignant behavior only ICD-O-3 behavior code of /3 (125,961 total cases for the 5-year period, for an annual average of 25,192 cases).

In comparison to other common cancers in the CiNA data set, brain and other CNS tumors are the sixth most commonly occurring tumor when using the CBTRUS definition (Figure 1, Panel A). Of the 481,650 cases of primary brain tumors, 137,617 (28.6%) were malignant (ICD-O-3 behavior code /3) while 344,033 (71.4%) were nonmalignant (ICD-O-3 behavior code /0 or /1) (Figure 1, Panel B). The overall AAAIR of brain and other CNS tumors was 24.44 per 100,000 (95% CI, 24.37–24.51). The incidence of malignant tumors only was 7.00 per 100,000 (95% CI, 6.96–7.04), while the incidence of nonmalignant tumors was 17.44 per 100,000 (95% CI, 17.38–17.50).

Brain and other CNS tumors occurred more frequently in females (AAAIR, 27.16 per 100,000; 95% CI, 27. 06–27.27) than in males (AAAIR, 21.55 per 100,000; 95% CI, 21.45–21.64). For malignant tumors only, incidence was higher in males than females (male AAAIR, 8.23 per 100,000; 95% CI, 8.18–8.29 compared to female AAAIR, 5.90 per 100,000; 95% CI, 5.85–5.95). The opposite was true in nonmalignant tumors, where incidence was nearly twice as high in females (female AAAIR, 21.26 per 100,000; 95% CI, 21.17–21.35) than in males (male AAAIR, 13.31 per 100,000; 95% CI, 13.24–13.39)

The most frequently occurring histopathology was meningioma (AAAIR, 9.09 per 100,000; 95% CI, 9.05–9.13), of which about 99% are nonmalignant (Table 2). Meningioma occurred more than twice as frequently in females (female AAAIR, 12.38 per 100,000; 95% CI, 12.31–12.45) than in males (male AAAIR, 5.40 per 100,000; 95% CI, 5.35–5.45). The second most frequently occurring histopathology was tumors of the pituitary (AAAIR, 4.28 per 100,000; 95% CI, 4.25–4.31), followed by glioblastoma (AAAIR, 3.20 per 100,000; 95% CI, 3.18–3.22). The predominantly nonmalignant tumors of the pituitary occurred at much higher rates in females than in males (female AAAIR, 4.83 per 100,000; 95% CI, 4.78–4.88 as compared to male AAAIR, 3.81 per 100,000; 95% CI, 3.77–3.86). Glioblastoma occurred more commonly in males than in females (male AAAIR, 4.00 per 100,000; 95% CI, 3.96–4.04 as compared to female AAAIR, 2.50 per 100,000; 95% CI, 2.48–2.54).

Incidence estimates by age group are shown in Table 4. Overall, brain and CNS tumors are most common in those older than 40 years (≥40 years AAAIR, 44.11 per 100,000; 95% CI, 43.97–44.25, as compared to 0–14 years AAAIR, 5.81 per 100,000; 95% CI, 5.73–5.90, and 15–39 years AAAIR, 11.84 per 100,000; 95% CI, 11.75–11.93). The most frequently occurring histopathology in children aged 0–14 years was pilocytic astrocytoma (AAAIR, 1.11 per 100,000; 95% CI, 1.07–1.15), while in adolescents and young adults aged 15–39 years, it was tumors of the pituitary (AAAIR, 4.03 per 100,000; 95% CI, 3.98–4.08). In older adults (≥40 years), meningioma was reported as the most frequently occurring histopathology (AAAIR, 19.49 per 100,000; 95% CI, 19.40–19.59).

Discussion

CBTRUS produces annual statistical reports using this CBTRUS histopathology grouping scheme, the most recent of which was based on 445,792 cases diagnosed from 2015–2019.[6] This includes data from 51 central cancer registries in the United States only (50 states and Washington DC; about 100% of the United States with the exception of Nevada from the years 2018–2019), and is therefore smaller than the CiNA data, which includes Canadian registries. The results of this initial evaluation show that, while case counts vary due to differences in included central cancer registries, the overall incidence patterns for primary brain and other CNS tumors are the same within the CBTRUS analytic data set and the NAACCR CiNA data set. Observed overall patterns, incidence by sex, and incidence by age are very similar to those estimated using the CBTRUS data set. Variations may occur when using the recode variable in data sets that may include lower proportions of nonmalignant tumors or variations in the proportion of cases abstracted from radiographic imaging, which is associated with increased detection (and therefore higher incidence) of nonmalignant brain and other CNS tumors. Care should be taken when generating and interpreting statistics for these tumors, particularly concerning the underlying site definition used, as these vary by major reporting organizations.

Incidence rates for brain and other CNS tumors vary significantly by histopathology, sex, age, and other demographic factors, including race/ethnicity and geography. The CBTRUS variable provides valuable information for comparing the burden of brain and other CNS tumors by these and other groups, as the etiology and diagnostic patterns of these histopathologies vary significantly. Use of this recode is useful not only for epidemiologic comparisons across populations, but also for investigating potential data collection or diagnostic differences across regions that may affect overall incidence rates of brain and other CNS tumors. While evaluation of brain and other CNS tumors by behavior is common, we encourage the use of the CBTRUS variable across behaviors to fully estimate the burden due to tumors of the brain and other CNS in populations of interest.

Based on the results of the initial evaluation of the CiNA version of the CBTRUS variable, NAACCR has updated the recode to select cases based on the CBTRUS

Table 2. Frequency of Cases Assigned to Categories in the CBTRUS and SEER Brain and CNS Histopathologic Recode Schemes, 2015–2019 NAACCR CiNA

		CBTRUS brain and CNS recode																												
		Diffuse astrocytoma	Anaplastic astrocytoma	Glioblastoma	Oligodendroglioma	Anaplastic oligodendroglioma	Oligoastrocytic tumors	Pilocytic astrocytoma	Unique astrocytoma variants	Ependymal tumors	Glioma malignant, NOS	Other neuroepithelial tumors	Neuronal and mixed neuronal-glial tumors	Choroid plexus tumors	Tumors of the pineal region	Embryonal tumors	Nerve sheath tumors	Other tumors of cranial and paraspinal nerves	Meningioma	Mesenchymal tumors	Primary melanocytic lesions	Lymphoma	Other hematopoietic neoplasms	Germ cell tumors	Tumors of the pituitary	Craniopharyngioma	Hemangioma	Neoplasm, unspecified	All other	Brain/CNS not categorized[a]	TOTAL
SEER brain and CNS recode	1. Malignant brain/CNS	9,653	7,518	67,795	4,041	2,160	595	5,537	572	4,106	9,060	66	257	124	<16	3,423	233	0	1,766	814	95	0	0	582	<16	<16	0	7,441	56	55	125,961
	1.1 Glioma	9,653	7,518	67,795	4,041	2,160	595	5,537	572	4,106	9,060	66	<16	0	<16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	111,115
	1.1.1 Diffuse astrocytoma and anaplastic astrocytoma	8,045	7,518	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15,563
	1.1.2 Glioblastoma	1,417	0	67,795	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	69,212
	1.1.3 Diffuse midline glioma, H3 K27M-mutant	0	0	0	0	0	0	0	0	0	412	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	412
	1.1.4 Oligodendroglioma	0	0	0	4,041	2,143	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6,184
	1.1.5 Oligoastrocytoma	0	0	0	0	0	595	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	595
	1.1.6 Other astrocytic tumors	0	0	0	0	0	0	5,537	559	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6,096
	1.1.7 Astroblastoma	0	0	0	0	0	0	0	0	0	0	61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	61
	1.1.8 Ependymal tumors	0	0	0	0	0	0	0	0	4,067	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4,067
	1.1.9 Glioma, unspecified	0	0	0	0	0	0	0	0	0	8,648	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8,648
	1.1.10 Other	191	0	0	0	17	0	0	<16	39	0	<16	<16	0	<16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	277
	1.2 Embryonal tumors	0	0	0	0	0	0	0	0	0	0	0	39	0	0	3,423	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3,462
	1.3 Meningiomas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,737	0	0	0	0	0	0	0	0	0	0	0	1,737
	1.4 Choroid plexus tumors	0	0	0	0	0	0	0	0	0	0	0	0	124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	124
	1.5 Neuronal and mixed neuronal-glial tumors	0	0	0	0	0	0	0	0	0	0	0	174	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	174
	1.6 Other malignant brain/CNS	0	0	0	0	0	0	0	0	0	0	0	33	0	<16	0	233	0	29	814	95	0	0	582	<16	<16	0	7,441	56	55	9,349
	2. Nonmalignant brain/CNS	23	<16	0	<16	<16	0	147	422	3,347	452	45	4,717	755	<16	<16	39,474	35	184,736	6,230	57	0	0	489	<16	405	4,372	12,010	146	95	257,988
	2.1 Meningiomas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	183,748	0	0	0	0	0	0	0	0	0	0	0	183,748
	2.2 Tumors of the cranial and paraspinal nerves	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38,737	0	0	0	0	0	0	0	0	0	0	0	0	0	38,737
	2.3 Ependymal tumors	0	0	0	0	0	0	0	0	3,178	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3,178
	2.4 Choroid plexus tumors	0	0	0	0	0	0	0	0	0	0	0	0	755	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	755
	2.5 Neuronal and mixed neuronal-glial tumors	0	0	0	0	0	0	0	0	0	0	0	4,250	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4,250
	2.6 Mesenchymal, nonmeningothelial tumors	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5,550	0	0	0	0	0	0	0	0	0	0	5,550
	2.7 Other astrocytic tumors	0	0	0	0	0	0	0	362	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	362
	2.8 Other gliomas	0	0	0	0	0	0	0	52	0	0	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97
	2.9 Other nonmalignant brain/CNS	23	<16	0	<16	<16	0	147	<16	169	452	0	467	0	<16	<16	737	35	988	680	57	0	0	489	<16	405	4,372	12,010	146	95	21,311
	3. Malignant tumors of the pineal region	0	0	0	0	0	0	0	0	0	0	0	0	0	478	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	478
	4. Nonmalignant tumors of the pineal region	0	0	0	0	0	0	0	0	0	0	0	0	0	260	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	260
	5. All other cancer types	<16	<16	32	0	<16	0	22	<16	<16	49	0	831	<16	87	25	<16	0	22	59	<16	9,210	49	698	79,996	2,977	<16	2,789	<16	66	96,963
	TOTAL	9,687	7,526	67,827	4,048	2,164	595	5,706	995	7,460	9,561	111	5,805	880	836	3,453	39,714	35	186,524	7,103	161	9,210	49	1,769	80,004	3,388	4,377	22,240	206	216	

CBTRUS, Central Brain Tumor Registry of the United States; CNS, central nervous system; ICD-O-3, *International Classification of Diseases for Oncology*, 3rd edition; NAACCR CiNA, North American Association of Central Cancer Registries Cancer in North America; CNS, central nervous system; SEER, Surveillance, Epidemiology, and End Results. Counts and rates are not presented when fewer than 16 cases were reported for the specific category. The suppressed cases are included in the counts and rates for totals. [a]Includes ICD-O-3 codes fitting site criteria that are not included in CBTRUS histopathology classification.

Table 3. Five-Year Total, Annual Average Total[a], and Average Annual Age-Adjusted Incidence Rates[b] with 95% Confidence Intervals for All Brain and Other Central Nervous System Tumors by Major Histopathology Groupings, Histopathology, Behavior, and Sex, 2015-2019 NAACCR CiNA

Histopathology	*Total*			*Male*			*Female*		
	5-Year Total	Annual Average	Rate (95% CI)	5-Year Total	Annual Average	Rate (95% CI)	5-Year Total	Annual Average	Rate (95% CI)
Diffuse astrocytic and oligodendroglial tumors	91,847	18,369	4.50 (4.47–4.53)	52,901	10,580	5.49 (5.44–5.54)	38,946	7,789	3.64 (3.60–3.68)
Diffuse astrocytoma	9,687	1,937	0.52 (0.51–0.53)	5,456	1,091	0.60 (0.59–0.62)	4,231	846	0.45 (0.43–0.46)
Anaplastic astrocytoma	7,526	1,505	0.40 (0.39–0.41)	4,120	824	0.45 (0.44–0.47)	3,406	681	0.35 (0.34–0.36)
Glioblastoma	67,827	13,565	3.20 (3.18–3.22)	39,549	7,910	4.00 (3.96–4.04)	28,278	5,656	2.50 (2.48–2.54)
Oligodendroglioma	4,048	810	0.23 (0.23–0.24)	2,247	449	0.26 (0.25–0.27)	1,801	360	0.20 (0.20–0.21)
Anaplastic oligodendroglioma	2,164	433	0.12 (0.11–0.12)	1,209	242	0.14 (0.13–0.14)	955	191	0.10 (0.09–0.11)
Oligoastrocytic tumors	595	119	0.03 (0.03–0.04)	320	64	0.04 (0.03–0.04)	275	55	0.03 (0.03–0.03)
Other astrocytic tumors	6,701	1,340	0.42 (0.41–0.43)	3,545	709	0.44 (0.42–0.45)	3,156	631	0.40 (0.39–0.42)
Pilocytic astrocytoma	5,706	1,141	0.36 (0.35–0.37)	3,003	601	0.37 (0.36–0.39)	2,703	541	0.35 (0.33–0.36)
Unique astrocytoma variants	995	199	0.06 (0.06–0.06)	542	108	0.07 (0.06–0.07)	453	91	0.06 (0.05–0.06)
Malignant	*573*	*115*	*0.03 (0.03–0.04)*	*296*	*59*	*0.04 (0.03–0.04)*	*277*	*55*	*0.03 (0.03–0.04)*
Nonmalignant	*422*	*84*	*0.03 (0.02–0.03)*	*246*	*49*	*0.03 (0.03–0.03)*	*176*	*35*	*0.02 (0.02–0.03)*
Ependymal tumors	7,460	1,492	0.42 (0.41–0.43)	4,302	860	0.49 (0.47–0.50)	3,158	632	0.35 (0.34–0.36)
Malignant	*4,110*	*822*	*0.23 (0.23–0.24)*	*2,250*	*450*	*0.26 (0.25–0.27)*	*1,860*	*372*	*0.21 (0.20–0.22)*
Nonmalignant	*3,350*	*670*	*0.18 (0.18–0.19)*	*2,052*	*410*	*0.23 (0.22–0.24)*	*1,298*	*260*	*0.14 (0.13–0.15)*
Other gliomas	9,672	1,934	0.55 (0.54–0.56)	4,909	982	0.57 (0.56–0.59)	4,763	953	0.53 (0.51–0.54)
Glioma malignant, NOS	9,561	1,912	0.54 (0.53–0.55)	4,865	973	0.57 (0.55–0.58)	4,696	939	0.52 (0.50–0.54)
Other neuroepithelial tumors	111	22	0.01 (0.01–0.01)	44	9	0.01 (0.00–0.01)	67	13	0.01 (0.01–0.01)
Neuronal and mixed neuronal-glial tumors	5,805	1,161	0.34 (0.33–0.35)	3,154	631	0.37 (0.36–0.38)	2,651	530	0.31 (0.30–0.32)
Malignant	*1,066*	*213*	*0.06 (0.05–0.06)*	*588*	*118*	*0.06 (0.06–0.07)*	*478*	*96*	*0.05 (0.05–0.05)*
Nonmalignant	*4,739*	*948*	*0.28 (0.28–0.29)*	*2,566*	*513*	*0.31 (0.29–0.32)*	*2,173*	*435*	*0.26 (0.25–0.27)*

Table 3, cont. Five-Year Total, Annual Average Total[a], and Average Annual Age-Adjusted Incidence Rates[b] with 95% Confidence Intervals for All Brain and Other Central Nervous System Tumors by Major Histopathology Groupings, Histopathology, Behavior, and Sex, 2015-2019 NAACCR CiNA

Histopathology	*Total*			*Male*			*Female*		
	5-Year Total	Annual Average	Rate (95% CI)	5-Year Total	Annual Average	Rate (95% CI)	5-Year Total	Annual Average	Rate (95% CI)
Choroid plexus tumors	880	176	0.05 (0.05–0.06)	441	88	0.05 (0.05–0.06)	439	88	0.05 (0.05–0.06)
Malignant	*125*	*25*	*0.01 (0.01–0.01)*	*74*	*15*	*0.01 (0.01–0.01)*	*51*	*10*	*0.01 (0.00–0.01)*
Nonmalignant	*755*	*151*	*0.04 (0.04–0.05)*	*367*	*73*	*0.04 (0.04–0.05)*	*388*	*78*	*0.04 (0.04–0.05)*
Tumors of the pineal region	d	167	0.05 (0.04–0.05)	352	70	0.04 (0.04–0.05)	484	97	0.06 (0.05–0.06)
Malignant	*508*	*102*	*0.03 (0.03–0.03)*	*243*	*49*	*0.03 (0.02–0.03)*	*265*	*53*	*0.03 (0.03–0.04)*
Nonmalignant	*328*	*66*	*0.02 (0.02–0.02)*	*109*	*22*	*0.01 (0.01–0.01)*	*219*	*44*	*0.02 (0.02–0.03)*
Embryonal tumors	3,453	691	0.22 (0.21–0.23)	2,086	417	0.26 (0.25–0.27)	1,367	273	0.18 (0.17–0.19)
Tumors of cranial and paraspinal nerves	39,749	7,950	2.02 (2.00–2.04)	19,099	3,820	2.02 (1.99–2.05)	20,650	4,130	2.03 (2.00–2.06)
Nerve sheath tumors	39,714	7,943	2.02 (2.00–2.04)	19,078	3,816	2.01 (1.99–2.04)	20,636	4,127	2.03 (2.00–2.05)
Malignant	*233*	*47*	*0.01 (0.01–0.01)*	*118*	*24*	*0.01 (0.01–0.02)*	*115*	*23*	*0.01 (0.01–0.01)*
Nonmalignant	*39,481*	*7,896*	*2.00 (1.98–2.02)*	*18,960*	*3,792*	*2.00 (1.97–2.03)*	*20,521*	*4,104*	*2.01 (1.99–2.04)*
Other tumors of cranial and paraspinal nerves	35	7	0.00 (0.00–0.00)	21	4	0.00 (0.00–0.00)	14	3	0.00 (0.00–0.00)
Tumors of meninges	193,788	38,758	9.48 (9.44–9.52)	54,414	10,883	5.80 (5.76–5.86)	139,374	27,875	12.76 (12.69–12.83)
Meningioma	186,524	37,305	9.09 (9.05–9.13)	50,719	10,144	5.40 (5.35–5.45)	135,805	27,161	12.38 (12.31–12.45)
Malignant	*1,768*	*354*	*0.09 (0.08–0.09)*	*823*	*165*	*0.09 (0.08–0.09)*	*945*	*189*	*0.09 (0.08–0.09)*
Nonmalignant	*184,756*	*36,951*	*9.00 (8.96–9.04)*	*49,896*	*9,979*	*5.31 (5.26–5.36)*	*134,860*	*26,972*	*12.29 (12.22–12.36)*
Mesenchymal tumors	7,103	1,421	0.38 (0.37–0.39)	3,603	721	0.40 (0.38–0.41)	3,500	700	0.37 (0.36–0.39)
Malignant	*851*	*170*	*0.05 (0.04–0.05)*	*452*	*90*	*0.05 (0.05–0.06)*	*399*	*80*	*0.04 (0.04–0.05)*
Nonmalignant	*6,252*	*1,250*	*0.34 (0.33–0.35)*	*3,151*	*630*	*0.35 (0.33–0.36)*	*3,101*	*620*	*0.33 (0.32–0.34)*
Primary melanocytic lesions	161	32	0.01 (0.01–0.01)	92	18	0.01 0.01–0.01)	69	14	0.01 (0.01–0.01)
Lymphomas and hematopoietic neoplasms	9,259	1,852	0.44 (0.43–0.45)	4,764	953	0.49 (0.48–0.51)	4,495	899	0.40 (0.39–0.41)
Lymphoma	9,210	1,842	0.44 (0.43–0.45)	4,736	947	0.49 (0.48–0.50)	4,474	895	0.40 (0.39–0.41)
Other hematopoietic neoplasms	49	10	0.00 (0.00–0.00)	28	6	0.00 (0.00–0.00)	21	4	0.00 (0.00–0.00)

Table 3, cont. Five-Year Total, Annual Average Total[a], and Average Annual Age-Adjusted Incidence Rates[b] with 95% Confidence Intervals for All Brain and Other Central Nervous System Tumors by Major Histopathology Groupings, Histopathology, Behavior, and Sex, 2015-2019 NAACCR CiNA

Histopathology	*Total*			*Male*			*Female*		
	5-Year Total	Annual Average	Rate (95% CI)	5-Year Total	Annual Average	Rate (95% CI)	5-Year Total	Annual Average	Rate (95% CI)
Germ cell tumors	1,769	354	0.11 (0.10–0.11)	1,226	245	0.15 (0.14–0.16)	543	109	0.07 (0.06–0.07)
Malignant	*1,229*	*246*	*0.08 (0.07–0.08)*	*938*	*188*	*0.11 (0.11–0.12)*	*291*	*58*	*0.04 (0.03–0.04)*
Nonmalignant	*540*	*108*	*0.03 (0.03–0.04)*	*288*	*58*	*0.03 (0.03–0.04)*	*252*	*50*	*0.03 (0.03–0.03)*
Tumors of sellar region	83,392	16,678	4.47 (4.44–4.50)	37,348	7,470	4.01 (3.97–4.05)	46,044	9,209	5.01 (4.96–5.06)
Tumors of the pituitary	80,004	16,001	4.28 (4.25–4.31)	35,593	7,119	3.81 (3.77–3.86)	44,411	8,882	4.83 (4.78–4.88)
Malignant	*118*	*24*	*0.01 (0.00–0.01)*						
Nonmalignant	*79,886*	*15,977*	*4.28 (4.25–4.31)*						
Craniopharyngioma	3,388	678	0.19 (0.18–0.19)	1,755	351	0.20 (0.19–0.21)	1,633	327	0.18 (0.17–0.19)
Unclassified tumors	26,973	5,395	1.37 (1.35–1.39)	12,213	2,443	1.36 (1.34–1.39)	14,760	2,952	1.39 (1.36–1.41)
Hemangioma	4,377	875	0.24 (0.23–0.24)	2,047	409	0.23 (0.22–0.24)	2,330	466	0.25 (0.24–0.26)
Neoplasm, unspecified	22,240	4,448	1.11 (1.10–1.13)	9,971	1,994	1.11 (1.09–1.13)	12,269	2,454	1.12 (1.10–1.14)
Malignant	*7,584*	*1,517*	*0.37 (0.36–0.38)*	*3,693*	*739*	*0.41 (0.40–0.43)*	*3,891*	*778*	*0.33 (0.32–0.34)*
Nonmalignant	*14,656*	*2,931*	*0.74 (0.73–0.76)*	*6,278*	*1,256*	*0.70 (0.68–0.71)*	*8,378*	*1,676*	*0.79 (0.77–0.81)*
All other	206	41	0.01 (0.01–0.01)	126	25	0.01 (0.01–0.02)	80	16	0.01 (0.01–0.01)
Brain/CNS not categorized[c]	150	30	0.01 (0.01–0.01)	69	14	0.01 (0.01–0.01)	81	16	0.01 (0.01–0.01)
Total[d]	481,650	96,330	24.44 (24.37–24.51)	200,781	40,156	21.55 (21.45-21.64)	280,869	56,174	27.16 (27.06–27.27)
Malignant	*137,617*	*27,523*	*7.00 (6.96–7.04)*	*76,982*	*15,396*	*8.23 (8.18–8.29)*	*60,635*	*12,127*	*5.90 (5.85–5.95)*
Nonmalignant	*344,033*	*68,807*	*17.44 (17.38–17.50)*	*123,799*	*24,760*	*13.31 (13.24-13.39)*	*220,234*	*44,047*	*21.26 (21.17–21.35)*

ICD-O-3, *International Classification of Diseases for Oncology*, 3rd edition; NOS, not otherwise specified.

[a]Annual average cases are calculated by dividing the five-year total by five

[b]Rates are per 100,000 and are age-adjusted to the 2000 US standard population.

[c]Includes ICD-O-3 codes fitting site criteria that are not included in CBTRUS histopathology classification

[d]Refers to all brain tumors including histopathologies not presented in this table.

Counts and rates are not presented when fewer than 16 cases were reported for the specific category. The suppressed cases are included in the counts and rates for totals.

Table 4. Five-Year Total, Annual Average Total,[a] and Average Annual Age-Adjusted Incidence Rates[b] with 95% CIs for All Brain and Other Central Nervous System Tumors by Major Histopathology Groupings, Histopathology, Behavior, and Age Group, 2015–2019 NAACCR CiNA

Histopathology	*0–14 y*			*15–39 y*			*≥40 y*		
	5-year total	Annual average	95% CI	5-year total	Annual average	95% CI	5-year total	Annual average	95% CI
Diffuse astrocytic and oligodendroglial tumors	1,394	279	0.46 (0.43–0.48)	10,737	2,147	1.84 (1.81–1.88)	79,598	15,920	8.71 (8.65–8.78)
Diffuse astrocytoma	630	126	0.21 (0.19–0.22)	2,901	580	0.49 (0.47–0.51)	6,113	1,223	0.70 (0.68–0.72)
Anaplastic astrocytoma	222	44	0.07 (0.06–0.08)	2,189	438	0.37 (0.36–0.39)	5,097	1,019	0.59 (0.57–0.60)
Glioblastoma	448	90	0.15 (0.13–0.16)	3,282	656	0.57 (0.55–0.59)	64,045	12,809	6.89 (6.83–6.94)
Oligodendroglioma	62	12	0.02 (0.02–0.03)	1,578	316	0.27 (0.26–0.29)	2,404	481	0.31 (0.29–0.32)
Anaplastic oligodendroglioma	–	–	–	–	–	–	1,574	315	0.19 (0.18–0.20)
Oligoastrocytic tumors	–	–	–	–	–	–	365	73	0.04 (0.04–0.05)
Other astrocytic tumors	3,760	752	1.23 (1.19–1.27)	1,982	396	0.34 (0.32–0.35)	821	164	0.10 (0.10–0.11)
Pilocytic astrocytoma	3,395	679	1.11 (1.07–1.15)	1,566	313	0.27 (0.25–0.28)	637	127	0.08 (0.07–0.09)
Unique astrocytoma variants	365	73	0.12 (0.11–0.13)	416	83	0.07 (0.06–0.08)	184	37	0.02 (0.02–0.03)
Malignant	–	–	–	*281*	*56*	*0.05 (0.04–0.05)*	–	–	–
Nonmalignant	–	–	–	*135*	*27*	*0.02 (0.02–0.03)*	–	–	–
Ependymal tumors	931	186	0.30 (0.28–0.32)	2,069	414	0.36 (0.34–0.37)	4,396	879	0.52 (0.51–0.54)
Malignant	*812*	*162*	*0.26 (0.25–0.28)*	*1,060*	*212*	*0.18 (0.17–0.19)*	*2,182*	*436*	*0.26 (0.25–0.27)*
Nonmalignant	*119*	*24*	*0.04 (0.03–0.05)*	*1,009*	*202*	*0.17 (0.16–0.18)*	*2,214*	*443*	*0.26 (0.25–0.27)*
Other gliomas	2,761	552	0.90 (0.87–0.94)	2,085	417	0.35 (0.34–0.37)	4,730	946	0.54 (0.53–0.56)
Glioma malignant, NOS	2,739	548	0.90 (0.86–0.93)	2,045	409	0.35 (0.33–0.36)	4,682	936	0.54 (0.52–0.55)
Other neuroepithelial tumors	22	4	0.01 (0.00–0.01)	40	8	0.01 (0.00–0.01)	48	10	0.01 (0.00–0.01)
Neuronal and mixed neuronal-glial tumors	1,391	278	0.46 (0.43–0.48)	2,353	471	0.40 (0.38–0.41)	1,968	394	0.24 (0.23–0.25)
Malignant	*86*	*17*	*0.03 (0.02–0.03)*	*210*	*42*	*0.04 (0.03–0.04)*	*758*	*152*	*0.09 (0.08–0.09)*
Nonmalignant	*1,305*	*261*	*0.43 (0.40–0.45)*	*2,143*	*429*	*0.36 (0.35–0.38)*	*1,210*	*242*	*0.15 (0.14–0.16)*

Table 4, cont. Five-Year Total, Annual Average Total,[a] and Average Annual Age-Adjusted Incidence Rates[b] with 95% CIs for All Brain and Other Central Nervous System Tumors by Major Histopathology Groupings, Histopathology, Behavior, and Age Group, 2015–2019 NAACCR CiNA

Histopathology	*0–14 y*			*15–39 y*			*≥40 y*		
	5-year total	**Annual average**	**95% CI**	**5-year total**	**Annual average**	**95% CI**	**5-year total**	**Annual average**	**95% CI**
Choroid plexus tumors	252	50	0.08 (0.07–0.09)	220	44	0.04 (0.03–0.04)	282	56	0.03 (0.03–0.04)
Malignant	*62*	*12*	*0.02 (0.02–0.03)*	–	–	–	–	–	–
Nonmalignant	*190*	*38*	*0.06 (0.05–0.07)*	–	–	–	–	–	–
Tumors of the pineal region	150	30	0.05 (0.04–0.06)	315	63	0.05 (0.05–0.06)	363	73	0.04 (0.04–0.05)
Malignant	*131*	*26*	*0.04 (0.04–0.05)*	*191*	*38*	*0.03 (0.03–0.04)*	*181*	*36*	*0.02 (0.02–0.03)*
Nonmalignant	*19*	*4*	*0.01 (0.00–0.01)*	*124*	*25*	*0.02 (0.02–0.03)*	*182*	*36*	*0.02 (0.02–0.03)*
Embryonal tumors	2,070	414	0.68 (0.65–0.71)	820	164	0.14 (0.13–0.15)	303	61	0.04 (0.03–0.04)
Tumors of cranial and paraspinal nerves	713	143	0.23 (0.22–0.25)	6,024	1,205	1.05 (1.02–1.07)	32,985	6,597	3.71 (3.67–3.75)
Nerve sheath tumors	713	143	0.23 (0.22–0.25)	6,017	1,203	1.05 (1.02–1.07)	32,957	6,591	3.71 (3.67–3.75)
Malignant	–	–	–	–	–	–	*164*	*33*	*0.02 (0.02–0.02)*
Nonmalignant	–	–	–	–	–	–	*32,793*	*6,559*	*3.69 (3.65–3.73)*
Other tumors of cranial and paraspinal nerves	–	–	–	–	–	–	28	6	0.00 (0.00–0.00)
Tumors of meninges	756	151	0.25 (0.23–0.27)	12,626	2,525	2.24 (2.20–2.28)	180,209	36,042	20.04 (19.94–20.13)
Meningioma	317	63	0.10 (0.09–0.12)	10,699	2,140	1.91 (1.88–1.95)	175,491	35,098	19.49 (19.40–19.59)
Malignant	–	–	–	–	–	–	*1,626*	*325*	*0.18 (0.17–0.19)*
Nonmalignant	–	–	–	–	–	–	*173,865*	*34,773*	*19.31 (19.22–19.41)*
Mesenchymal tumors	429	86	0.14 (0.13–0.15)	1,898	380	0.32 (0.31–0.34)	4,596	919	0.53 (0.52–0.55)
Malignant	*72*	*14*	*0.02 (0.02–0.03)*	*184*	*37*	*0.03 (0.03–0.04)*	*590*	*118*	*0.07 (0.06–0.07)*
Nonmalignant	*357*	*71*	*0.12 (0.10–0.13)*	*1,714*	*343*	*0.29 (0.28–0.31)*	*4,006*	*801*	*0.46 (0.45–0.48)*
Primary melanocytic lesions	10	2	0.00 (0.00–0.01)	29	6	0.00 (0.00–0.01)	122	24	0.01 (0.01–0.02)
Lymphomas and hematopoietic neoplasms	87	17	0.03 (0.02–0.04)	580	116	0.10 (0.09–0.11)	8,588	1,718	0.94 (0.92–0.96)

Table 4, cont. Five-Year Total, Annual Average Total,[a] and Average Annual Age-Adjusted Incidence Rates[b] with 95% CIs for All Brain and Other Central Nervous System Tumors by Major Histopathology Groupings, Histopathology, Behavior, and Age Group, 2015–2019 NAACCR CiNA

Histopathology	*0–14 y*			*15–39 y*			*≥40 y*		
	5-year total	Annual average	95% CI	5-year total	Annual average	95% CI	5-year total	Annual average	95% CI
Lymphoma	–	–	–	–	–	–	8,545	1,709	0.93 (0.91–0.95)
Other hematopoietic neoplasms	–	–	–	–	–	–	43	9	0.00 (0.00–0.01)
Germ cell tumors	680	136	0.22 (0.21–0.24)	792	158	0.13 (0.12–0.14)	215	43	0.03 (0.02–0.03)
Malignant	*513*	*103*	*0.17 (0.15–0.18)*	*635*	*127*	*0.11 (0.10–0.11)*	–	–	–
Nonmalignant	*167*	*33*	*0.05 (0.05–0.06)*	*157*	*31*	*0.03 (0.02–0.03)*	–	–	–
Tumors of sellar region	1,811	362	0.59 (0.57–0.62)	24,408	4,882	4.16 (4.11–4.21)	57,157	11,431	6.68 (6.62–6.73)
Tumors of the pituitary	1,110	222	0.36 (0.34–0.39)	23,663	4,733	4.03 (3.98–4.08)	55,222	11,044	6.45 (6.40–6.51)
Malignant	–	–	–	–	–	–	*98*	*20*	*0.01 (0.01–0.01)*
Nonmalignant	–	–	–	–	–	–	*55,124*	*11,025*	*6.44 (6.39–6.50)*
Craniopharyngioma	701	140	0.23 (0.21–0.25)	745	149	0.13 (0.12–0.14)	1,935	387	0.22 (0.21–0.23)
Unclassified tumors	1,014	203	0.33 (0.31–0.35)	3,748	750	0.64 (0.62–0.66)	22,096	4,419	2.48 (2.45–2.51)
Hemangioma	278	56	0.09 (0.08–0.10)	1,207	241	0.21 (0.19–0.22)	2,876	575	0.34 (0.32–0.35)
Neoplasm, unspecified	698	140	0.23 (0.21–0.25)	2,491	498	0.43 (0.41–0.45)	18,970	3,794	2.11 (2.08–2.15)
Malignant	*146*	*29*	*0.05 (0.04–0.06)*	*380*	*76*	*0.07 (0.06–0.07)*	*7,029*	*1,406*	*0.77 (0.76–0.79)*
Nonmalignant	*552*	*110*	*0.18 (0.17–0.20)*	*2,111*	*422*	*0.36 (0.35–0.38)*	*11,941*	*2,388*	*1.34 (1.31–1.36)*
All other	28	6	0.01 (0.01–0.01)	23	5	0.00 (0.00–0.01)	142	28	0.02 (0.01–0.02)
Brain/CNS not categorized[c]	–	–	–	–	–	–	108	22	0.01 (0.01–0.01)
Total[d]	17,773	3,555	5.81 (5.73–5.90)	68,780	13,756	11.84 (11.75-11.93)	393,753	78,751	44.11 (43.97–44.25)
Malignant	*11,421*	*2,284*	*3.73 (3.67–3.80)*	*18,841*	*3,768*	*3.22 (3.18–3.27)*	*106,578*	*21,316*	*11.74 (11.67–11.82)*
Nonmalignant	*6,352*	*1,270*	*2.08 (2.03–2.13)*	*49,939*	*9,988*	*8.62 (8.54–8.69)*	*287,175*	*57,435*	*32.36 (32.24–32.49)*

CBTRUS, Central Brain Tumor Registry of the United States; CNS, central nervous system; NAACCR CiNA, North American Association of Central Cancer Registries Cancer in North America; NOS, not otherwise specified. Annual average cases are calculated by dividing the 5-year total by 5. Counts and rates are not presented when fewer than 16 cases were reported for the specific category. The suppressed cases are included in the counts and rates for totals.

[a]Annual average cases are calculated by dividing the five-year total by five

[b]Rates are per 100,000 and are age-adjusted to the 2000 US standard population.

[c]Includes ICD-O-3 codes fitting site criteria that are not included in CBTRUS histopathology classification

[d]Refers to all brain tumors including histopathologies not presented in this table.

site definition. This variable, *CBTRUS Histology Recode,* will be included in the June 2023 release of all CiNA data sets and will be available for cancer registries and researchers to calculate on their own data via File*Pro. NAACCR and CBTRUS will collaborate annually on how best to update the variable if coding or groupings change. This variable is intended to be a more dynamic, clinically driven definition for all primary brain and other CNS tumors. Therefore, users of the recode must be mindful that the CBTRUS reporting definition for all primary brain and other CNS tumors is based on a clinical model supported by the various revisions of the *WHO Classification of Tumours of the Central Nervous System*. Currently, this is reflective of the 2016 WHO CNS classification. As a result, care is needed when using this classification for analyses of data collected prior to 2016, particularly for histopathologic categories (eg, gliomas) that have changed substantially in classification over time.

We encourage cancer registries and researchers to use this variable to ensure more detailed, comprehensive, and clinically relevant assessment of the burden of this important group of tumors.

References

1. Schoenberg BS, Christine BW, Whisnant JP. The descriptive epidemiology of primary intracranial neoplasms: the Connecticut experience. *Am J Epidemiol*. 1976;104(5):499-510. doi:10.1093/oxfordjournals.aje.a112322
2. Kruchko C, Ostrom QT, Gittleman H, Barnholtz-Sloan JS. The CBTRUS story: providing accurate population-based statistics on brain and other central nervous system tumors for everyone. *Neuro Oncol*. 2018;20(3):295-298. doi:10.1093/neuonc/noy006
3. Davis FG, Malinski N, Haenszel W, et al. Primary brain tumor incidence rates in four United States regions, 1985–1989: A Pilot Study. *Neuroepidemiology*. 1996;15(2):103-112. doi:10.1159/000109895
4. Benign Brain Tumor Cancer Registries Amendment Act, 107th Cong. § 260 (2002). **https://www.gpo.gov/fdsys/pkg/PLAW-107publ260/pdf/PLAW-107publ260.pdf**
5. McCarthy BJ, Surawicz T, Bruner JM, Kruchko C, Davis F. Consensus conference on brain tumor definition for registration. *Neuro Oncol*. 2000;4(2):134-45. doi:10.1093/neuonc/4.2.134 1920658
6. Ostrom QT, Price M, Neff C, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019. *Neuro Oncol*. 2022;24(suppl 5):v1-v95. doi:10.1093/neuonc/noab200
7. Louis DN OH, Wiestler OD, Cavanee WK, ed. *WHO Classification of Tumours of the Central Nervous System*. International Agency for Research on Cancer; 2016.
8. Iorgulescu JB, Sun C, Neff C, et al. Molecular biomarker-defined brain tumors: epidemiology, validity, and completeness in the United States. *Neuro Oncol*. 2022;24(11):1989-2000. doi:10.1093/neuonc/noac113
9. Waite KA, Cioffi G, Kruchko C, et al. Aligning the Central Brain Tumor Registry of the United States (CBTRUS) histology groupings with current definitions. *Neurooncol Pract*. 2022;9(4):317-327. doi:10.1093/nop/npac025
10. Forjaz G, Barnholtz-Sloan JS, Kruchko C, et al. An updated histology recode for the analysis of primary malignant and nonmalignant brain and other central nervous system tumors in the Surveillance, Epidemiology, and End Results Program. *Neuro Oncol Adv*. 2021;3(1):vdaa175. doi:10.1093/noajnl/vdaa175
11. Sherman R, Firth R, Kahl A, et al, eds. *Cancer in North America: 2015–2019. Volume One: Combined Cancer Incidence for the United States, Canada and North America*. North American Association of Central Cancer Registries, Inc; 2022. **https://www.naaccr.org/wp-content/uploads/2022/06/CiNA.2015-2019.v1.combined-incidence.pdf**
12. North American Association of Central Cancer Registries Inc. (NAACCR). SEER*Stat Database: NAACCR Incidence Data—CiNA Research Data, 1995-2019, for Expanded Races, Standard File, Sherman—CBTRUS Recode Evaluation (which includes data from CDC's National Program of Cancer Registries (NPCR), the Canadian Cancer Registry's Provincial and Territorial Registries, and the NCI's Surveillance, Epidemiology and End Results (SEER) Registries), certified by the North American Association of Central Cancer Registries (NAACCR) as meeting high-quality incidence data standards for the specified time periods, submitted December 2021. 2022.
13. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat software version 8.4.0. National Cancer Institute, DCCPS, Surveillance Research Program. Accessed May 27, 2022. **www.seer.cancer.gov/seerstat**

Original Article

Current and Emerging Informatics Initiatives Impactful to Cancer Registries

Dana E. Rollison, PhD[a]; Gary M. Levin, CTR[b]; Jeremy L. Warner, MD, MS, FAMIA, FASCO[c,d]; Rich Pinder[e]; Lori A. Havener, CTR[f]; Madhusmita Behera, PhD[g,h]; Andrew R. Post[i]; Rajan Gopalakrishnan[j]; Eric B. Durbin, DrPH, MS[k,l]

Abstract: Cancer surveillance at the population level is a highly labor-intensive process, with certified tumor registrars (CTRs) manually reviewing medical charts of cancer patients and entering information into local databases that are centrally merged and curated at state and national levels. Registries face considerable challenges in terms of constrained budgets, staffing shortages, and keeping pace with the evolving national and international data standards that are essential to cancer registration. Advanced informatics methods are needed to increase automation, reduce manual efforts, and to help address some of these challenges. The Cancer Informatics Advisory Group (CIAG) to the North American Association of Central Cancer Registries (NAACCR) board was established in 2019 to advise of external informatics activities and initiatives for long-term strategic planning. Reviewed here by the CIAG are current informatics initiatives that were either born out of the cancer registry field or have implications for expansion to cancer surveillance programs in the future. Several areas of notable activity are presented, including an overview of informatics initiatives and descriptions of 12 specific informatics projects with implications for cancer registries. Recommendations are also provided to the registry community for the continued tracking and impact of the projects and initiatives.

Key words: *automation, Cancer Informatics Advisory Group, cancer surveillance*

Introduction

Cancer surveillance at the population level is a highly labor-intensive process, with certified tumor registrars (CTRs) manually reviewing medical charts of cancer patients and entering information into local databases that are centrally merged and curated at state and national levels. This manual process results in considerable delays in the availability of data at the population level. Furthermore, expansion of data acquisition through cancer registries is expensive to scale, given the reliance on CTRs, who are difficult to recruit, train, and retain. Advanced informatics techniques that can semiautomate casefinding, data abstraction, aggregation, and curation are needed to expand the types of data captured in cancer registries to advance our understanding of population-level trends in cancer incidence, mortality, and health disparities and to measure the effectiveness of cancer prevention and screening programs. Methods that reduce manual labor and increase registry workflow efficiencies have the potential to improve the overall timeliness of data capture and publication. In addition, national and international data standards are essential for the continued evolution of cancer registration. The ability to capture current concepts in health care and oncology requires the cancer registry community to follow new and emerging standards and be prepared to adopt them when appropriate.

Reviewed here are current informatic initiatives that were either born out of the cancer registry field or have implications for expansion to cancer surveillance programs in the future. Areas of notable activity include electronic pathology (ePath) reporting, information extraction from electronic health records (EHRs), and emerging data standards. Prepared by the Cancer Informatics Advisory Group (CIAG) to the North American Association of Central Cancer Registries (NAACCR) board, the following review includes project summaries, links to additional project information, and assessments of the potential intersection with cancer surveillance. Specific recommendations to the registry community for the continued tracking and impact of the projects and initiatives are also provided.

Overview of Informatics Initiatives

Electronic Data Sources and Associated Information Retrieval

Pathology represents the most important data source for cancer surveillance, as most cancers (85%–90%) are diagnosed through microscopic examination of human tissue and often reported by a pathology laboratory. For over 2 decades, registries have developed and implemented standards and technologies to facilitate automated electronic reporting of pathology results, although the proportion of ePath reporting varies by registry. However, an ongoing challenge of ePath reporting is that the elements of most importance (site, histology, behavior, grade, and laterality) are embedded within the narrative text of the report.

[a]Moffitt Cancer Center, Tampa, Florida. [b]Florida Cancer Data System. [c]Lifespan Cancer Institute. [d]Vanderbilt University Medical Center, Nashville, Tennessee. [e]Los Angeles Cancer Surveillance Program, Los Angeles, California. [f]North American Association of Central Cancer Registries, Springfield, Illinois. [g]Winship Cancer Institute, Atlanta, Georgia. [h]Woodruff Health Sciences Center, Emory University, Atlanta, Georgia. [i]Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah. [j]Comprehensive Cancer Center, University of Chicago, Chicago, Illinois. [k]Kentucky Cancer Registry, Lexington, Kentucky. [l]Markey Cancer Center, University of Kentucky, Lexington, Kentucky.

Address correspondence to Eric B. Durbin, DrPH, MS. Email: ericd@kcr.uky.edu.

Structured ePath reporting or automated extraction of the key ePath elements has the greatest potential to improve the efficiency of registry abstracting and operations.

We report on several advances in ePath technologies and initiatives. The most extensive and successful effort to date has been the National Cancer Institute (NCI)–US Department of Energy (DOE) natural language processing (NLP) project. This collaborative project between the DOE and the NCI Surveillance, Epidemiology, and End Results (SEER) Program has harnessed the state-of-the-art computing facilities at the DOE to produce a deployable application programming interface (API) that automates the coding of the key elements, improves accuracy, and provides a confidence metric for the prediction. As of 2021, it is being deployed in SEER registries with plans to make it available to other registries soon. A similar project funded by NCI is cancer deep phenotype extraction for the cancer registry *(DeepPhe*CR),* which involves developing NLP methods to extract structured data from both ePath and EHR data sources. DeepPhe*CR is still experimental, and investigators are developing methods to extract data beyond the key elements, such as biomarker information. Neither of these projects includes the data transport mechanism between pathology laboratories and cancer registries. The National Program of Cancer Registries (NPCR)'s ePath reporting project is part of a broader effort sponsored by the Centers for Disease Control and Prevention (CDC) to implement secure reporting of ePath data from national laboratories to central cancer registries through the CDC's Association of Public Health Laboratories Informatics Messaging Services (AIMS) platform. The CDC–NPCR Data Modernization Initiative (DMI) extends beyond ePath reporting and is more broadly moving resources into a cloud environment to support a variety of cancer and noncancer surveillance activities. The goal of this initiative is to improve the efficiency of cancer reporting by allowing reporters to send data to a centralized data repository that will be directed to the appropriate state. The ultimate goal of the CDC's DMI is to get better, faster, and actionable insights for decision-making at all levels of public health. The CDC's Childhood Cancer Survivorship, Treatment, Access, and Research (STAR) Project is part of the same effort but is focused specifically on the early reporting of pediatric and young adult cancers. While the NPCR projects do not yet include structured reporting or NLP-derived coding, the NCI–DOE API has a good chance of serving as a complementary component since it is already in production. Alignment and collaboration across these complementary efforts could accelerate the overall advancement of the field.

While ePath reports provide a wealth of information, they are not always available, nor are they complete. EHRs continue to be a promising, but not yet realized, source of reliable, semiautomated cancer surveillance data. Fundamental challenges arise from an EHR data model that differs significantly from the cancer surveillance patient-tumor-treatment view that is essential for public health reporting and used in cancer prevention and control. EHR data are collected as chronological clinical tests, procedures, and events that are not necessarily tied to a particular cancer diagnosis and are often not discrete or structured in nature. Cancer diagnosis date, for example, is not a concept routinely captured in EHR problem lists, though it can likely be derived from the corpus of EHR documents for a patient. Several technologies and initiatives described below are underway that may prove beneficial to cancer surveillance. Fast Healthcare Interoperability Resources (FHIR) is a Health Level Seven International (HL7) technology supported by EHR vendors that facilitates APIs for the exchange of information between EHRs and entities such as registries. The Making EHR Data More Available for Research and Public Health (MedMorph) initiative has emerged from patient-centered outcomes research and leverages FHIR technology. It is not clear that FHIR and MedMorph address the patient-tumor-treatment challenge, but they may represent an opportunity for more reliable and structured EHR data extraction for registries. Digital Bridge is an industry- and vendor-centric initiative to facilitate data transport within existing EHR systems. FHIR is one of several transports supported by Digital Bridge. CDC-sponsored "connectathons" have successfully demonstrated a proof of concept for Digital Bridge transports, which are now being used for COVID-19 reporting. As these initiatives mature, population-based cancer registries should continue to identify ways to leverage these technologies to improve the efficiency and completeness of cancer case ascertainment and reporting.

A summary of the informatics initiatives presented, including brief descriptions, active dates, links, and recommended next steps for the registry community, is provided in Table 1.

Data Standards

Several emerging data standard initiatives may impact cancer surveillance through their potential to enable data harmonization across systems. *International Classification of Diseases, 11th revision (ICD-11)* is the next-generation coding system from the World Health Organization (WHO) that will eventually replace ICD-10. While released by the WHO on January 1, 2022, it is not clear when it will be implemented in the United States, which typically creates a "clinical modification (CM)" that can lag significantly (for example, ICD-10-CM lagged ICD-10 by approximately 23 years). Once ICD-11-CM is adopted in the United States and Canada, it will have significant implications for casefinding and cancer reporting due to considerable advancements in the semantic structure of the terminology. Minimal Common Oncology Data Elements (mCODE) emerged from the American Society of Clinical Oncology (ASCO) in collaboration with several industry partners. mCODE is a nonproprietary data model and a trial standard set of FHIR profiles designed to capture key elements pertinent to cancer patient clinical data. It is not yet clear how widely mCODE will be adopted, and the data model is likely to undergo additional evolution before becoming a normative standard. HemOnc, started in 2018 and funded by the NCI through at least 2027, is an ontology built to formally represent chemotherapy regimens and protocols.[1] In addition to basic relationships between regimen names and drug components, HemOnc captures the context of treatment

Table 1. Summary of Informatics Initiatives

Initiative	*Description*	*Active dates*	*Website*	*Next steps*
NCI–Department of Energy (DOE) Natural Language Processing (NLP) Project	Machine learning and natural language processing (NLP)-based automated extraction of key elements from ePath reports; deployed within SEER registries	2021–present	**https://surveillance.cancer.gov/research/nlp.html**	Share application programming interface (API) with other central registries, improve model accuracy, and expand elements such as biomarkers and reoccurrence
Deep Phenotype Extraction (DeepPhe)/ Cancer Deep Phenotype Extraction for the Cancer Registry (DeepPhe*CR)	NLP-based automated extraction of detailed patient phenotypic data from ePath reports and the EHR	2019–2025	**https://deepphe.github.io/software/**	Studies ongoing to demonstrate software utility
NPCR Electronic Pathology Reporting Project	Cloud-based informatics platform to facilitate secure reporting of cancer ePath reports from national laboratories to central cancer registries	2019–present	**https://www.cdc.gov/cancer/npcr/informatics/aerro/index.htm**	Develop structured reporting and NLP-derived coding; align with complimentary efforts (NCI-DOE NLP)
CDC/NPCR Data Modernization Initiative (DMI)	Multipronged initiative designed to advance real-time cancer reporting through data acquisition, data transmission, application requirements, cloud-based environments, and data standards	2019–present	**https://www.cdc.gov/surveillance/surveillance-data-strategies/data-IT-transformation.html**	Remain aware of DMI and participate in funding and deployment opportunities as they arise.
CDC's Childhood Cancer Survival, Treatment, Access, and Research (STAR) Project	An informatics system designed to facilitate reporting of pediatric and young adult cancers (PAYAC) to national cancer registries	2018–present	**https://www.cdc.gov/cancer/npcr/pediatric-young-adult-cancer.htm**	Partner with key agencies and NPCR exports to solicit feedback and best practices on informatics infrastructure; accelerate PAYAC surveillance efforts, research, and enrollment in clinical trials
Fast Healthcare Interoperability Resources (FHIR)	A Health Level Seven (HL7) technology supported by EHR vendors to facilitate interoperability formats for exchange of clinical and meta data between EHRs and entities (such as registries)	2012–present	**https://www.hl7.org/fhir/**	Remain aware of FHIR initiatives, particularly those supported by CDC/NPCR, NCI/SEER, and NAACCR. Participate in funding and deployment opportunities as they arise.
Making EHR Data More Available for Research and Public Health (MedMorph)	A patient-centered outcomes research (PCOR) initiative that aims to increase availability of EHR data through the HL7 FHIR mechanism	2020–present	**https://build.fhir.org/ig/HL7/fhir-medmorph/**	Three use cases are being developed and were submitted to the HL7 balloting processes in January 2023
International Classification of Diseases (ICD)-11	A coding framework for systematically recording human health conditions across the globe including more than 17,000 diagnostic categories	2022	**https://icd.who.int/en**	Pending adoption in United States and Canada to improve cancer case finding and reporting

Table 1, cont. Summary of Informatics Initiatives				
Initiative	*Description*	*Active dates*	*Website*	*Next steps*
Minimal Common Oncology Data Elements (mCODE)	A coding framework specific to cancer, including domains related to patient, laboratory/vital, disease, genomics, treatment, and outcome	2018–present	**https://confluence.hl7.org/display/COD/mCODE**	Adoption of mCODE by organizations is ongoing
NCI Cancer Data Standards Registry and Repository (caDSR)	A repository to manage and use data standards by providing shared standards in human and machine-readable contexts	1997–present	**https://datascience.cancer.gov/resources/metadata**	Development on next generation of csDSR to include minimal set of v18 and v21
Pathology, Radiology, Imaging, Signs, Symptoms, Medical oncology, bioMarkers (PRISSMM)	A common set of definitions to measure cancer disease progression including use of pathology, radiology, imaging, signs, symptoms, medical oncology, and biomarkers	2018–present	**https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8314138/**	Disease progression is a valuable concept not currently captured by cancer registries; future adoption of disease progression measures by NAACCR may involve the standards developed by PRISSMM

CDC, Centers for Disease Control and Prevention; EHR, electronic health record; ePath, electronic pathology; NAACCR, North American Association of Central Cancer Registries; NCI, National Cancer Institute; NPCR, National Program of Cancer Registries; SEER, Surveillance, Epidemiology, and End Results Program.

(eg, adjuvant, first-line metastatic) and temporal complexities. In collaboration with SEER under the auspices of the Childhood Cancer Data Initiative, HemOnc is proactively working towards increased integration with the registry community, as it would facilitate the capture of standard representations of complex chemotherapy protocols. The NCI Cancer Data Standards Registry and Repository (caDSR) and its associated applications help the oncology research community manage and use data standards by providing the shared standards in various human- and machine-readable contexts. Used primarily in research, the availability of NAACCR data elements in the caDSR may promote the use and uptake of registry data elements.

Of note, most data standards focus on cancer diagnosis and treatment without attempting to define measures of cancer outcomes, such as treatment response, disease recurrence, and progression. For many years, the clinical trial community has relied on radiographic parameters to define response and progression, typically using the Response Evaluation Criteria in Solid Tumors (RECIST) definitions. While these are still useful in many clinical trial settings, they are onerous to capture in routine clinical care and do not capture nonradiologic progression. Pathology, Radiology, Imaging, Signs, Symptoms, Medical oncology, bioMarkers (PRISSMM) is a set of instructions and definitions proposed for the measurement of disease progression that may be superior to RECIST as the result of its multiaxial approach to temporal status changes.[2] It has been developed and licensed by the Dana-Farber Cancer Institute, but the extent of uptake and use is not clear at this time. Many consider the lack of population-based data on treatment response, recurrence, and progression to be a major limitation of central registries, and the cancer registry community should closely track the uptake of PRISSMM and similar efforts with an eye towards future inclusion. Advocacy for reducing barriers caused by licensing fees and licensing restrictions may be required for broad uptake.

Descriptions of Specific Informatics Projects

Natural Language Processing (NLP) Project

The NCI–DOE NLP project is a collaborative effort involving the NCI SEER Program, the DOE, and a number of SEER registries. The objective is to enhance the utility of ePath reporting via automated coding of key elements such as site, histology, behavior, laterality, and grade from narrative text contained within unstructured ePath reports. Automated coding of ePath reports relies on machine-learning NLP techniques to derive predictions of the key elements from large volumes of previous ePath reports submitted to the participating SEER registries that have been linked to abstracted cases. The coded key elements in the abstracts serve as the reference standard for the machine learning. The DOE employs its scientific expertise in machine-learning NLP and massive computational resources with domain expertise from the SEER Program to build models that have shown very good performance in predicting correct values for the key elements, particularly for frequently occurring cancer sites such as lung, breast, colorectal, and others.[3-10] Collaborators have also developed a confidence measure that is calculated along with the predictions. The predictive models are being implemented in registry operations through an API that can be deployed within a secure registry setting. Because the models were developed using narrative text, there is the potential for protected health information to be embedded within the

underlying model components. Registries must therefore ensure appropriate security protections when deploying such models. The API accepts ePath reports as input and returns the predictions and confidence score as output. SEER has already deployed the models for its registries and there are plans for the API to be made available for other central registries. The collaborative work is ongoing towards improving the key element coding accuracy and incorporating additional elements such as recurrence and biomarkers.

More information is available at **https://surveillance.cancer.gov/research/nlp.html**.

*DeepPhe/DeepPhe*CR*

DeepPhe and DeepPhe*CR are related NCI-funded NLP projects. DeepPhe, started in 2014, tackles the entire EHR, whereas DeepPhe*CR, started in 2019, focuses on central cancer registry feeds. These NLP initiatives seek to extract precise patient phenotype information that is needed to advance translational cancer research, particularly to unravel the effects of genetic, epigenetic, and systems changes on tumor behavior and responsiveness. Examples of phenotypic variables in cancer include tumor morphology (eg, histopathologic diagnosis), comorbid conditions (eg, associated immune disease), laboratory findings (eg, gene amplification status), specific tumor behaviors (eg, metastasis), and response to treatment (eg, effect of a chemotherapeutic agent on a tumor). Current models for correlating EHR data with ~*omics* data largely ignore the clinical text, which remains one of the most important sources of phenotype information for cancer patients. Unlocking the value of clinical text has the potential to enable new insights about cancer initiation, progression, metastasis, and response to treatment. A diverse set of oncology studies led by accomplished translational investigators in breast cancer, melanoma, ovarian cancer, and colorectal cancer will demonstrate the utility of the DeepPhe software during the project period (2020–2025 for DeepPhe and 2019–2024 for DeepPhe*CR). DeepPhe software bridges novel methods to automate cancer deep phenotype extraction from clinical text with emerging standards in phenotype knowledge representation and NLP. This work is highly aligned with recent calls in the scientific literature to advance scalable and robust methods of extracting and representing phenotypes for precision medicine and translational research. Specific examples relevant for cancer surveillance include determining biomarker status for breast cancer patients and determining grade, histology, laterality, and other standard parameters for a variety of cancer types. Summaries of both projects are available on the NCI Reporter.[11,12] For more information, visit **https://deepphe.github.io/software/**.

Electronic Pathology (ePath) Reporting Project

This project is part of the DMI to implement the use of the AIMS platform to improve the public health agencies data exchange infrastructure between laboratories and public health agencies. Historically, laboratories have reported cancer pathology data by paper or nonstandard electronic formats (Excel, ASCII file, etc) to state cancer registries. Over the past decade, the CDC NPCR staff worked with national and large regional laboratories to implement the NAACCR volume V standard for reporting cancer pathology data to state cancer registries. Over 30 national and large regional laboratories have successfully created standard electronic messages and set up secure data exchange connectivity with each state cancer registry. For national and large regional laboratories, maintaining so many secure data exchange connections with public health agencies has created a burden for laboratories and public health agencies. The CDC's Center for Surveillance, Epidemiology, and Laboratory Services (CSELS) infectious disease electronic laboratory reporting (ELR) program successfully implemented standardized reporting from Quest Diagnostics using the AIMS platform. AIMS is an Amazon cloud platform that has served as infrastructure to receive, validate, and distribute ELR data from the laboratory to the appropriate state departments of health. The use of the AIMS platform has minimized the resources required by laboratories and public health agencies to exchange critical public health data. This project will provide laboratories and state cancer registries with access to the AIMS infrastructure and technical assistance to support real-time data exchange of standards-based cancer pathology data to meet mandated public health reporting to cancer registries. CDC does not receive or have access to the reported confidential patient data.

For more information, visit **https://www.cdc.gov/cancer/npcr/informatics/aerro/index.htm**.

Data Modernization Initiative (DMI)

The CDC/NPCR DMI goal is to modernize the field of cancer surveillance to achieve real-time cancer reporting. The NPCR and the CDC, as an agency, depend on high-quality, readily available data. Efficient electronic data exchange, interoperability, and data acquisition are of tantamount importance and a top priority of the DMI initiative. The initiative operates under 5 separate workgroups made up of funded central cancer registry participants. This is a one-year project to be completed by the end of the current cooperative agreement period. The 5 workgroups include:

- *ePath/Association of Public Health Laboratories activities workgroup.* The purpose of this workgroup is to onboard national, regional, and local laboratories to transmit cancer pathology and biomarker data to the AIMS platform for distribution to state cancer registries.
- *Self-service vendor/provider onboarding workgroup.* The purpose of this workgroup is to establish requirements for the development of a self-service onboarding application for vendors and providers to establish connectivity to the cancer surveillance cloud-based computing platform for electronic cancer case reporting.
- *Electronic Mapping, Reporting, and Coding (eMaRC) Plus cloud requirement and testing workgroup.* eMaRC Plus is an application used by cancer registries that processes pathology and clinical reports from pathology laboratories and physician offices using HL7 standard document formats. It is designed to create registry

abstracts from these reports. The purpose of this workgroup is to review eMaRC Plus functionality, test, and provide feedback on eMaRC Plus development to be used in the cloud and gather information on possible future enhancements to eMaRC Plus, both in the short and long term.

- *Web Plus cloud requirement and testing workgroup.* Web Plus provides online cancer data capture and abstracting functionality for central registries. The purpose of this workgroup is to review Web Plus functionality, test, and provide feedback on Web Plus development to be used in the cloud and gather information on possible future enhancements to Web Plus, both in the short and long term.
- *Data governance in the cloud workgroup.* The purpose of this workgroup is to establish data access standards/roles, security standards, honest broker relationships, best practices, business service agreements, and other cloud data governance topics.

More information is available at **https://www.cdc.gov/surveillance/surveillance-data-strategies/data-IT-transformation.html**.

Childhood Cancer Survivorship, Treatment, Access, and Research (STAR) Project

In June 2018, the Childhood Cancer STAR Act was signed into law to specifically help address the burden of childhood cancer. The law asked the CDC to expand capacity within the NPCR to help cancer registries collect and make the data on pediatric cancer cases available within weeks of diagnosis. Increasing the speed of reporting (within 30 days of diagnosis) increases the value of the data beyond routine cancer surveillance. Electronic reporting must be the standard by which this is achieved.

The aim of the Childhood Cancer STAR Project is to improve the timeliness of pediatric, adolescent, and young adult (ages 0–29 years) cancers through enhanced informatics systems and central cancer registry capacity across all registries through more effective and efficient electronic reporting. The goal is for facilities to have greater knowledge of gaps in pediatric, adolescent, and young adult cancers and that reporting facilities will report cases more rapidly.

On September 30, 2019, the CDC awarded a contract to Tanaq Support Services to execute the deliverables of the Childhood Cancer STAR Project. Tanaq is an 8(a) Alaskan Native Corporation with experienced program management. Tanaq focuses on delivering high-quality services in management consulting; software development; data management and analysis; epidemiology and technical assistance; health communication and social media; graphic design; web content development; technical writing; and training. Through the Childhood Cancer STAR Project, Tanaq will develop an informatics system to build on existing registry infrastructure effectively and efficiently for rapid case ascertainment of pediatric, adolescent, and young adult cancer incidences. Moreover, Tanaq will support select grantees to effectively pilot the newly designed informatics system to improve pediatric data collection.

The Childhood Cancer STAR project will also augment relationships with major partners, such as the St. Baldrick's Foundation, the Alliance for Childhood Cancer, and the Children's Oncology Group, by convening an expert panel to solicit feedback regarding linkages, data sharing, and data use. In addition, Tanaq will convene NPCR expert roundtables to discuss and implement best practices into the developed informatics system. The STAR project aims to improve patient enrollment into clinical trials through improved timeliness of the data collection, to provide more immediate use to researchers and public health professionals for program planning and identifying gaps in cancer prevalence, and to improve electronic cancer pathology and biomarker reporting for the NPCR overall. Additionally, the project will provide pediatric cancer data that public health practitioners and others can use to develop effective interventions. More timely data will also help reveal inequities in distribution of disease, treatment access, and other issues that can then be addressed.

For more information, visit **https://www.cdc.gov/cancer/npcr/pediatric-young-adult-cancer.htm**.

Fast Healthcare Interoperability Resources (FHIR)

FHIR is the fourth major effort of Health Level Seven International (HL7) to create a widely adopted interoperability format for the exchange of clinical data and metadata. To date, the most successful HL7 effort has been the second, called *HL7 V2,* which is still widely used by laboratory and other data exchange systems. HL7 V3 was widely seen as too complex, and only small portions were ever adopted, such as the Consolidated Clinical Document Architecture. The idea behind FHIR is to package chunks of clinical data into resources and assemble these using profiles, employing a widely adopted Representational State Transfer (REST) design.[13] Registries may benefit from FHIR as the result of applications that may more easily interface with EHRs to access or extract information pertinent to cancer registration. The specifications are maintained on a living web page: **https://www.hl7.org/fhir/**.

Making EHR Data More Available for Research and Public Health (MedMorph)

MedMorph is an initiative sponsored by the US Department of Health and Human Services Office of the Secretary Patient-Centered Outcomes Research Trust Fund. It began in 2020 with a goal to make EHR data more available to public health professionals and researchers through the HL7 FHIR transport mechanism. The project uses health data and exchange standards to develop and implement an interoperable solution to enable access to clinical data, incorporating these data into cancer registries' data systems. Three use cases are being developed currently, with submission to the HL7 balloting process scheduled for January 2023. CDC is leading the work on the cancer use case and is incorporating NAACCR community involvement in the process. CDC hosts webinars to ensure that the process will be guided by adequate representation from the cancer registry community. The MedMorph implementation guide is a complex specification document defining

how MedMorph integrates with the HL7 FHIR ecosystem. NAACCR is a member of the HL7 standards organization.

More information is available via the links below:

- **https://build.fhir.org/ig/HL7/fhir-medmorph/**
- **https://build.fhir.org/ig/HL7/fhir-medmorph/background.html**
- **https://www.cdc.gov/csels/phio/making_rapid_data_exchange_reality.html**
- **https://www.cdc.gov/nchs/nvss/modernization/pdf/fhir-implimentation-guidance-checklist.pdf**
- **https://www.cdc.gov/csels/phio/exchanging-data-efficiently.html**

International Classification of Diseases, 11th revision (ICD-11)

The International Classification of Diseases (ICD) has provided a coding framework for systematically recording human health conditions across the globe for more than a century. The first ICD edition was adopted by WHO in 1900 with subsequent revisions issued every 10 to 15 years. The 10th edition currently in use internationally was approved by the WHO in 1989, although its use in the United States didn't begin until 1999.

ICD codes are essential for calculating condition-specific incidence and mortality statistics for comparisons across countries and for tracking trends over time. ICD codes are also often used for administrative purposes, such as reimbursement and resource allocation, clinical research, and disease surveillance. More than 2 decades after the implementation of ICD-10, work began on the 11th revision, ICD-11, which was adopted by WHO on May 25, 2019, and went into effect on January 1, 2022. However, it is unclear when the United States will implement ICD-11.

ICD-11 was developed through a multidisciplinary collaboration of hundreds of clinicians, statisticians, coders, and information technology professionals across 90 countries, incorporating up-to-date and highly relevant clinical content in a fully electronic form through a multilingual REST API. ICD-11 includes more than 17,000 diagnostic categories and over 100,000 diagnostic terms, with new categories added for diseases of the immune system and other noncancer areas of focus. Of particular importance for cancer surveillance, the *International Classification of Disease for Oncology, Third Edition (ICD-O-3)* can be derived from ICD-11, with links to ICD-O-3 embedded within ICD-11. Furthermore, ICD-11 includes an addendum for newly defined "extension codes" to be used in conjunction with conventional "stem codes," whereby tumor, node, and metastasis (TNM) stage and histology can be coded along with primary site.[14] If fully adopted in the United States, the level of specificity available through ICD-11 stem and extension codes could improve the utility of billing diagnosis codes for timely case ascertainment. Furthermore, linkage of data to other sources may also be improved, with easier cross-walking between ICD-11 and other terminologies, such as Systematized Nomenclature of Medicine (SNOMED).

More information is available via the links below:

- **https://icd.who.int/en**
- **https://www.cdc.gov/nchs/data/icd/ICD-11-WHOV-CM-2018-V3.pdf**

Minimal Common Oncology Data Elements (mCODE)

mCODE is a project that originally emerged from ASCO and is now also supported by several industry partners. The mCODE specification is organized by 6 high-level domains: patient, laboratory/vital, disease, genomics, treatment, and outcome. The initial release of mCODE in 2018 consisted of 23 FHIR profiles composed of 90 data elements. The most recent iteration has expanded to 30 profiles. A manuscript describing the promise of mCODE to improve cancer data interoperability was published in 2020.[15] Adoption of the mCODE standards by EHR vendors may benefit registries by providing electronic access to patient data pertinent to registries. For example, registries could access or obtain standardized patient information through queries or FHIR interfaces. Additional information is also available from HL7 international[16]: **https://confluence.hl7.org/display/COD/mCODE**

Cancer Data Standards Registry and Repository (caDSR)

The NCI fosters the shared metadata standards for all cancer data that link together semantic meaning and data value. These standards promote sharing, reuse, and aggregation of cancer data among repositories. To assist with this interoperability, the NCI established the caDSR. The caDSR and its associated applications help the oncology research community manage and use data standards by providing the shared standards in various human and machine-readable contexts. The metadata content development team supports the oncology research community in developing and maintaining harmonized and standardized metadata for oncology research.[17]

The caDSR includes NAACCR standards 11 through 16. In November 2020, 2 representatives from the NCI Center for Biomedical Informatics and Information Technology (CBIIT) content team presented to the CIAG. At that time, it was recommended that NAACCR identify a minimal set for v18 and v21 in the fall of 2021. The NCI CBIIT content team were working on a next generation of caDSR to launch in the summer of 2021 and would not have the staff to work on the NAACCR content. Another consideration is the time and effort it takes NAACCR to review and verify the updated information.

More information on these efforts is available here: **https://datascience.cancer.gov/resources/metadata**

Pathology, Radiology, Imaging, Signs, Symptoms, Medical oncology, bioMarkers (PRISSMM)

PRISSMM is a systematic approach to the measurement of change in cancer status over time. Unlike RECIST, which depends entirely on morphometric measurements, PRISSMM takes a multiaxial approach to temporal status changes. The axes are pathology, radiology, imaging (somewhat distinct from radiology), signs, symptoms, medical oncology, and biomarkers. While there is no manuscript

specifically describing the system and the protocol, there have been several manuscripts evaluating the output, such as this article that we recently published: **https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8314138/**. The PRISSMM model is currently owned and licensed by Dana-Farber Cancer Institute; it is free for academic use.

Conclusion

The initiatives described in this report span academic, industry, and government domains, yet alignment and coordination of efforts will be required to advance the cancer surveillance field overall. New informatics advances in data acquisition, integration, linkage, and curation proven to work in a focused area or population should be tested in other areas/populations with the ultimate goal of scaling these advances for population-based cancer surveillance. Furthermore, processes are needed to drive consensus in data standards across source systems and domains to advance interoperability and data aggregation at the population level. Taken together, these solutions are needed to augment current cancer registries with additional clinical information of emerging relevance, thus advancing our understanding of cancer etiology and outcomes and driving innovation in cancer control.

Acknowledgements

The authors wish to acknowledge the ongoing support received by the CIAG of the NAACCR from Karen M. Knight, Chief of Registry Development, NAACCR. We also acknowledge the content contributions pertaining to the NPCR ePath reporting project by Sandy Jones from the CDC.

References

1. Warner JL, Dymshyts D, Reich CG, et al. HemOnc: a new standard vocabulary for chemotherapy regimen representation in the OMOP common data model. *J Biomed Inform*. 2019;96:103239.
2. Griffith SD, Miksad RA, Calkins G, et al. Characterizing the feasibility and performance of real-world tumor progression end points and their association with overall survival in a large advanced non-small-cell lung cancer data set. *JCO Clin Cancer Inform*. 2019;3:1-13.
3. De Angeli K, Gao S, Blanchard A, et al. Using ensembles and distillation to optimize the deployment of deep learning models for the classification of electronic cancer pathology reports. *JAMIA Open*. 2022;5(3):ooac075
4. Yoon H-J, Peluso A, Durbin EB, et al. Automatic information extraction from childhood cancer pathology reports. *JAMIA Open*. 2022;5(2):ooac049.
5. Blanchard AE, Gao S, Yoon HJ, et al. A keyword-enhanced approach to handle class imbalance in clinical text classification. *IEEE J Biomed Health Inform*. 2022;26(6):2796-2803.
6. De Angeli K, Gao S, Danciu I, et al. Class imbalance in out-of-distribution datasets: improving the robustness of the TextCNN for the classification of rare cancer types. *J Biomed Inform*. 2022;125:103957.
7. Yoon HJ, Stanley C, Christian JB, et al. Optimal vocabulary selection approaches for privacy-preserving deep NLP model training for information extraction and cancer epidemiology. *Cancer Biomark*. 2022;33:185-198.
8. De Angeli K, Gao S, Alawad M, et al. Deep active learning for classifying cancer pathology reports. *BMC Bioinformatics*. 2021;22(1):113.
9. Yoon H-J, Klasky HB, Gounley JP, et al. Accelerated training of bootstrap aggregation-based deep information extraction systems from cancer pathology reports. *J Biomed Inform*. 2020;110:103564.
10. Gao S, Alawad M, Schaefferkoetter N, et al. Using case-level context to classify cancer pathology reports. *PLoS One*. 2020;15(5):e0232840
11. Cancer deep phenotype extraction from electronic medical records. National Institutes of Health RePORTER website. Accessed September 27, 2022. **https://reporter.nih.gov/search/xNjEWpEsmkuSLUEMDSmRMw/project-details/10058470**
12. Natural language processing platform for cancer surveillance. National Institutes of Health RePORTER website. Accessed September 27, 2022. **https://reporter.nih.gov/search/xNjEWpEsmkuSLUEMDSmRMw/project-details/9980862**
13. Fielding RT. *Architectural Styles and the Design of Network-Based Software Architectures*. University of California; 2000.
14. Drösler SE, Weber S, Chute CG. ICD-11 extension codes support detailed clinical abstraction and comprehensive classification. *BMC Med Inform Decis Mak*. 2021;21:278.
15. Osterman TJ, Terry M, Miller RS. Improving cancer data interoperability: the promise of the Minimal Common Oncology Data Elements (mCODE) initiative. *JCO Clin Cancer Inform*. 2020;4:993-1001.
16. Minimal Common Oncology Data Elements (mCODE) implementation guide. Health Level Seven International website. Accessed September 27, 2022. **https://hl7.org/fhir/us/mcode/**
17. Metadata Services for Cancer data. National Cancer Institute's Center for Biomedical Informatics & Information Technology website. Updated September 20, 2021. Accessed September 27, 2022. **https://datascience.cancer.gov/resources/metadata**

Original Article

Identifying Factors Associated with Loss to Follow-up Among Patients Reported to the New York State Cancer Registry

Baozhen Qiao, PhD[a]; April A. Austin, MPH[a]; Maria J. Schymura, PhD[a]; Xiuling Zhang, PhD[a]; Colleen G. Sherman, MPH[a]

Abstract: Background: State cancer registries in the United States are data sources for estimating population-based cancer survival. However, the completeness of patient follow-up can affect the accuracy of survival estimates. Like many registries, the New York State Cancer Registry (NYSCR) conducts patient follow-up largely through linkages with other data sources. Even after expending great effort on linkages, a small proportion of patients remain lost to follow-up (LTFU). In this study, we identified factors that are associated with the likelihood of LTFU in the NYSCR. Methods: First primary cancers (sequence number, 00 or 01 and excluding death-certificate- and autopsy-only cases) diagnosed during 2000–2018 among New York State residents were selected for study. All patients were followed through December 31, 2018. Based on each patient's vital status and last contact date, follow-up status was categorized into 2 groups: patients LTFU and patients not LTFU. Patients LTFU were examined by demographic and tumor characteristics. Multivariate logistic regression analyses were performed to evaluate the association between demographic/tumor characteristics and likelihood of LTFU. For patients LTFU, the timing of LTFU (within 1 year, 1 to <5 years, 5 to <10 years, or >10 years) was further described. LTFU rates within 5 years after cancer diagnosis were also examined. Results: Among 1,797,228 patients, 74,722 were LTFU prior to December 31, 2018, representing 4.2% of all patients and 7.6% of alive patients. About 60% of LTFU occurred within 1 year after cancer diagnosis. Compared to the reference group, logistic regression analyses indicated that patients LTFU were more likely to be female, Black, Asian/Pacific Islander (API), Hispanic, foreign born, insured by Medicaid, uninsured, aged <20 years, and living in New York City or metropolitan counties. Cases reported by laboratories only and physician offices also had a higher likelihood of LTFU. Similar patterns and effects were identified when evaluating 5-year LTFU. Conclusion: Identifying factors associated with patient LTFU is important for cancer registries to improve follow-up data. We found that LTFU is not random; rather, certain patient groups have higher LTFU rates than others. For registries that conduct follow-up through linkages, it is critical to collect high-quality and complete demographic data, especially for females, children, the foreign born, and minority race/ethnicity groups.

Key words: *cancer registries, follow-up, loss to follow-up, New York State Cancer Registry, survival*

Introduction

State cancer registries in the United States are data sources for estimating population-based cancer survival.[1-6] However, the completeness of patient follow-up in registries can affect the accuracy of survival estimates.[7-12] Follow-up procedures vary among registries and are dependent on funding agency requirements and available resources.[11] Currently, other than death ascertainment, there are no other follow-up standards required for registries funded by the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR).[6,13] However, registries funded by the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program need to meet certain contractual follow-up standards. For instance, SEER registries are expected to meet follow-up goals of at least 90% for patients diagnosed under age 65 year and 95% for patients diagnosed at age 65 years or older.[14]

In general, there are 2 types of patient follow-up used by registries: active follow-up or passive follow-up.[3,12] Active follow-up requires contact with patients, physicians, family members, or other informants. Because active follow-up is costly and time-consuming, registries are unable to implement the method on a large scale. Instead, most registries in the United States use the passive method to update patient follow-up information.[3,7,8,12] The passive method largely relies on linkage with other data sources, such as state vital records, the National Death Index (NDI), hospital discharge data, insurance claims, and Social Security Administration (SSA) data. Previous studies reported that complete death ascertainment is especially important for registries using the passive follow-up method.[6-8,15]

The New York State Cancer Registry (NYSCR), funded by the NPCR since 1995 and by the SEER program since 2018, is one of the largest registries in the nation, collecting

[a]New York State Department of Health, Bureau of Cancer Epidemiology, Albany, NY

Address correspondence to Baozhen Qiao, PhD, 150 Broadway, Suite 361, Albany NY 12204. Telephone: (518) 474-2255.
Email: Baozhen.qiao@health.ny.gov.

This project was funded in part by the Centers for Disease Control and Prevention (CDC)'s National Program of Cancer Registries through cooperative agreement 6NU58DP006309 awarded to the New York State Department of Health and by the National Cancer Institute, National Institutes of Health (NIH), Department of Health and Human Services, under Contract 75N91018D00005. The contents are solely the responsibility of the New York State Department of Health and do not necessarily represent the official views of the CDC or NIH.

data on more than 120,000 newly diagnosed cancer cases each year. As a NPCR registry, the NYSCR has a long history of linking cancer data with state vital records and the NDI to identify deaths. After joining the SEER program, to meet the follow-up standards, the NYSCR has expanded its follow-up activities.[16,17] Like many other SEER registries, the NYSCR primarily uses the passive method to conduct patient follow-up. Even though the NYSCR has been able to meet the SEER program's follow-up standards, a small proportion of patients remain lost to follow-up (LTFU). In this study, we identified factors that are associated with patient LTFU in the NYSCR. With this knowledge, we hope to improve our follow-up information even further.

Materials and Methods

Study Population

First malignant primary cancers (sequence number 00 or 01) diagnosed during 2000–2018 among New York State residents were selected for this study. Death-certificate-only (DCO) and autopsy-only cases were excluded. All patients were followed through December 31, 2018, which is the follow-up date that was established and evaluated by the SEER program for the November 2021 data submission. Based on each patient's vital status and date of last contact, follow-up status was categorized into 2 groups: *patients not LTFU,* including those deceased or those alive with a date of last contact of December 31, 2018, or later, and *patients LTFU,* including those not known to be deceased (referred to as "alive") with a date of last contact prior to December 31, 2018.

The demographic factors that were evaluated include sex (male, female), age at cancer diagnosis (<20, 20–64, ≥65 years), race (White, Black, American Indian/Alaska Native, API, unknown), ethnicity (non-Hispanic, Hispanic), birth place (US born, foreign born, unknown), New York State region (New York City, New York State excluding New York City), census tract poverty level (assigned based on address at diagnosis: 0% to <5%, 5% to <10%, 10% to <20%, 20% to 100%, unknown), rural–urban county status (metropolitan, nonmetropolitan) and insurance status at diagnosis (insured, any Medicaid, uninsured, unknown). The tumor factors that were evaluated include year of cancer diagnosis (2000–2004, 2005–2009, 2010–2014, 2015–2018), tumor behavior (in situ, invasive), stage at diagnosis (in situ/local, regional, distant, unknown), type of reporting source (hospital inpatient, radiation treatment or oncology center, laboratory only, physician office, other hospital outpatient/surgery center), diagnostic confirmation (microscopically confirmed, clinical diagnosis only, unknown), sequence number (only 1 primary, first of multiple primaries), and cancer site group assigned according to disease survival rates (cancers with the best survival, including prostate, testis, thyroid, and melanoma of the skin; cancers with good survival, including breast, cervix, uterus, bladder, and kidney; cancers with bad survival, including oral cavity, colorectal, larynx, ovary, myeloma, leukemia, and others not listed in another group; cancers with the worst survival, including brain/central nervous system, esophagus, stomach, lung and bronchus, liver, and pancreatic).

Statistical Analyses

The number of patients LTFU was tabulated by demographic and tumor characteristics, and the associated percentage was calculated in 2 ways. First, we calculated the percentages by including all patients (both deceased and alive) in the denominators, which is how follow-up rates are evaluated by the SEER program.[14] Because deceased patients were automatically considered as having complete follow-up and because certain factors investigated in this study (eg, age or stage at cancer diagnosis) could affect survival, we also used a second approach: calculating the percentages of patients LTFU with only alive patients included in the denominators to remove the effect of disease fatality differences. For patients who were LTFU, the timing of LTFU (within 1 year, 1 to <5 years, 5 to <10 years, or ≥10 years after cancer diagnosis) was further examined by demographic and tumor characteristics. To evaluate the association between demographic/tumor characteristics and likelihood of LTFU, multivariate logistic regression analyses were performed.

Because 5-year survival rates are commonly used for measuring disease prognosis among cancer patients, we also conducted secondary analyses to evaluate rates of LTFU within 5 years after cancer diagnosis. These analyses were restricted to patients diagnosed during 2000–2013 to ensure that all patients had 5 complete years of follow-up. Vital status and follow-up status were recoded to correspond to the 5-year follow-up time. All statistically analyses were performed using SAS 9.4.

Results

Among 1,797,228 patients diagnosed 2000–2018 and included in this study, 74,722 were LTFU prior to December 31, 2018, representing 4.2% of all patients and 7.6% of alive patients. The detailed frequencies of LTFU by demographic and tumor characteristics are displayed in Table 1. When all patients were included in the analysis, higher percentages of LTFU were observed for the following demographic/tumor groups: female (4.5%), age < 20 years at cancer diagnosis (8.8%), API (12.3%), Hispanic (7.8%), foreign born (7.9%), uninsured (18.6%), living in New York City (6.5%) or metropolitan counties (4.4%), and tumors with in situ behavior (6.3%), diagnosed at an early stage (4.9%), reported by laboratory only (14.1%) or physician office (11.1%), being microscopically confirmed (4.3%), being the only primary cancer diagnosis (4.6%), or cancer with best prognosis (5.9%). As might be expected, when only alive patients were considered, the percentages of LTFU were always higher than when all patients were considered, but the general patterns of LTFU were consistent except for a few factors that were associated with disease prognosis. For example, the percentage of LTFU was higher for cancer cases diagnosed at distant stage, with only a clinical diagnosis, or with poor survival.

The multivariate logistic regression analyses (Table 2) indicated that all demographic/tumor factors evaluated were statistically significantly associated with the likelihood of LTFU. Notably, patients who were female, Black or API, Hispanic, foreign born, insured by Medicaid or uninsured,

Table 1. Frequency Distributions of Patients Who Were Lost to Follow-up (LTFU) Prior to December 31, 2018, or Within 5 Years After Cancer Diagnosis, By Demographic/Tumor Characteristics

Demographic/tumor characteristics	*Patients LTFU prior to December 31, 2018*[a]			*Patients LTFU within 5 y after cancer diagnosis*[b]		
	n	% Of all patients (n = 1,797,228)	% Of alive patients (n = 989,924)	n	% Of all patients (n = 1,304,137)	% Of alive patients (n = 799,687)
Total	74,722	4.2	7.6	37,714	2.9	4.7
Sex						
Male	33,024	3.8	7.3	15,932	2.5	4.2
Female	41,698	4.5	7.8	21,782	3.3	5.2
Age (y)						
<20	1,598	8.8	10.4	725	5.5	3.5
20–64	44,813	5.1	7.3	25,899	4.1	5.5
≥65	28,311	3.1	7.9	11,090	1.7	6.5
Race						
White	46,395	3.2	5.9	22,447	2.1	3.5
Black	13,115	5.1	9.7	7,278	4.0	6.9
American Indian/Alaska Native	80	3.0	5.1	43	2.4	3.9
Asian/Pacific Islander	10,115	12.3	18.4	5,344	10.2	15.1
Unknown	5,017	28.0	30.9	2,602	27.1	30.0
Ethnicity						
Non-Hispanic	61,808	3.8	6.9	30,259	2.5	4.2
Hispanic	12,914	7.8	13.0	7,455	6.6	10.2
Birth country						
US born	14,628	1.4	3.5	6,275	0.8	1.6
Foreign born	22,813	7.9	15.8	14,231	6.6	11.5
Unknown	37,281	7.9	8.8	17,208	5.8	6.3
Region						
New York City	43,350	6.5	11.7	23,658	4.9	8.0
New York State excluding New York City	31,372	2.8	5.1	14,056	1.7	2.8
Poverty level						
0% – <5%	16,520	3.5	6.0	7,911	2.2	3.4
5% – <10%	17,562	3.7	6.6	8,310	2.4	3.9
10% – <20%	20,362	4.3	7.9	10,093	3.0	5.1
20%–100%	19,839	5.4	10.5	11,094	4.2	7.4
Unknown	439	8.3	16.5	306	6.7	10.7
Rural–urban						
Metropolitan counties	72,099	4.4	7.9	36,694	3.1	5.0
Nonmetropolitan counties	2,623	1.8	3.5	1,020	0.9	1.6
Insurance status						
Insured	32,960	2.8	4.8	12,303	1.6	2.5
Any Medicaid	15,448	7.0	12.9	7,406	5.4	10.1
Uninsured	4,005	18.6	31.7	2,791	16.5	26.5
Unknown	22,309	5.6	13.4	15,214	4.1	7.1

Table 1, cont. Frequency Distributions of Patients Who Were Lost to Follow-up (LTFU) Prior to December 31, 2018, or Within 5 Years After Cancer Diagnosis, By Demographic/Tumor Characteristics

Demographic/tumor characteristics	*Patients LTFU prior to December 31, 2018*[a]			*Patients LTFU within 5 y after cancer diagnosis*[b]		
	n	% Of all patients (n = 1,797,228)	% Of alive patients (n = 989,924)	n	% Of all patients (n = 1,304,137)	% Of alive patients (n = 799,687)
Year of diagnosis						
2000–2004	15,446	3.5	9.9	13,352	3.0	5.2
2005–2009	16,721	3.5	7.3	11,142	2.4	3.8
2010–2014	20,895	4.3	7.1	13,220	3.4	5.3
2015–2018	21,660	5.5	7.0	-	-	-
Tumor behavior						
In situ	7,883	6.3	7.4	3,681	4.3	4.6
Invasive	66,839	4.0	7.6	34,033	2.8	4.7
Stage						
In Situ/local	43,822	4.9	6.7	21,479	3.4	4.0
Regional	12,328	3.6	7.0	6,857	2.8	4.9
Distant	9,184	2.3	8.7	4,972	1.7	6.5
Unknown	9,388	5.8	16.6	4,406	3.4	8.4
Type of reporting source						
Hospital inpatient	34,430	3.3	7.6	20,078	2.5	4.9
Radiation therapy or medical oncology center	7,058	4.1	5.5	2,913	2.4	3.0
Laboratory only	5,252	14.1	20.0	2,262	8.4	10.5
Physician office	5,681	11.1	14.2	1,569	5.3	6.8
Other hospital outpatient/ surgery center	22,301	4.6	6.6	10,892	3.3	4.4
Diagnostic confirmation						
Microscopically confirmed	72,677	4.3	7.5	36,516	3.0	4.7
Clinical diagnosis only	1,674	2.2	11.8	954	1.7	8.8
Unknown	371	2.3	13.3	244	1.7	6.3
Sequence number						
Only 1 primary	71,292	4.6	8.3	36,557	3.3	5.7
First of multiple primaries	3,430	1.4	2.5	1,157	0.6	0.7
Cancer site group[c]						
Cancers with best survival	24,522	5.9	7.8	10,852	3.6	4.2
Cancers with good survival	26,282	4.5	6.6	14,150	3.4	4.4
Cancers with bad survival	16,997	3.7	8.3	8,719	2.6	5.0
Cancers with worst survival	6,921	2.0	9.8	3,993	1.6	8.3

[a]Patients were diagnosed during 2000–2018 and followed through December 31, 2018. [b]Patients were diagnosed during 2000–2013 and followed for 5 years after cancer diagnosis. [c]Cancer site group was assigned according to disease survival rates. Cancers with best survival include prostate, testis, thyroid, and melanoma of the skin. Cancers with good survival include breast, cervix, uterus, bladder, and kidney. Cancers with bad survival include oral cavity, colorectal, larynx, ovary, myeloma, leukemia, and others not listed in another group. Cancers with worst survival include brain/central nervous system, esophagus, stomach, lung/bronchus, and liver/pancreatic.

Table 2. Multivariate Logistic Regression Analyses Evaluating Association Between Demographic/Tumor Characteristics and Likelihood of Loss to Follow-up (LTFU), Odd Ratios, and 95% CI

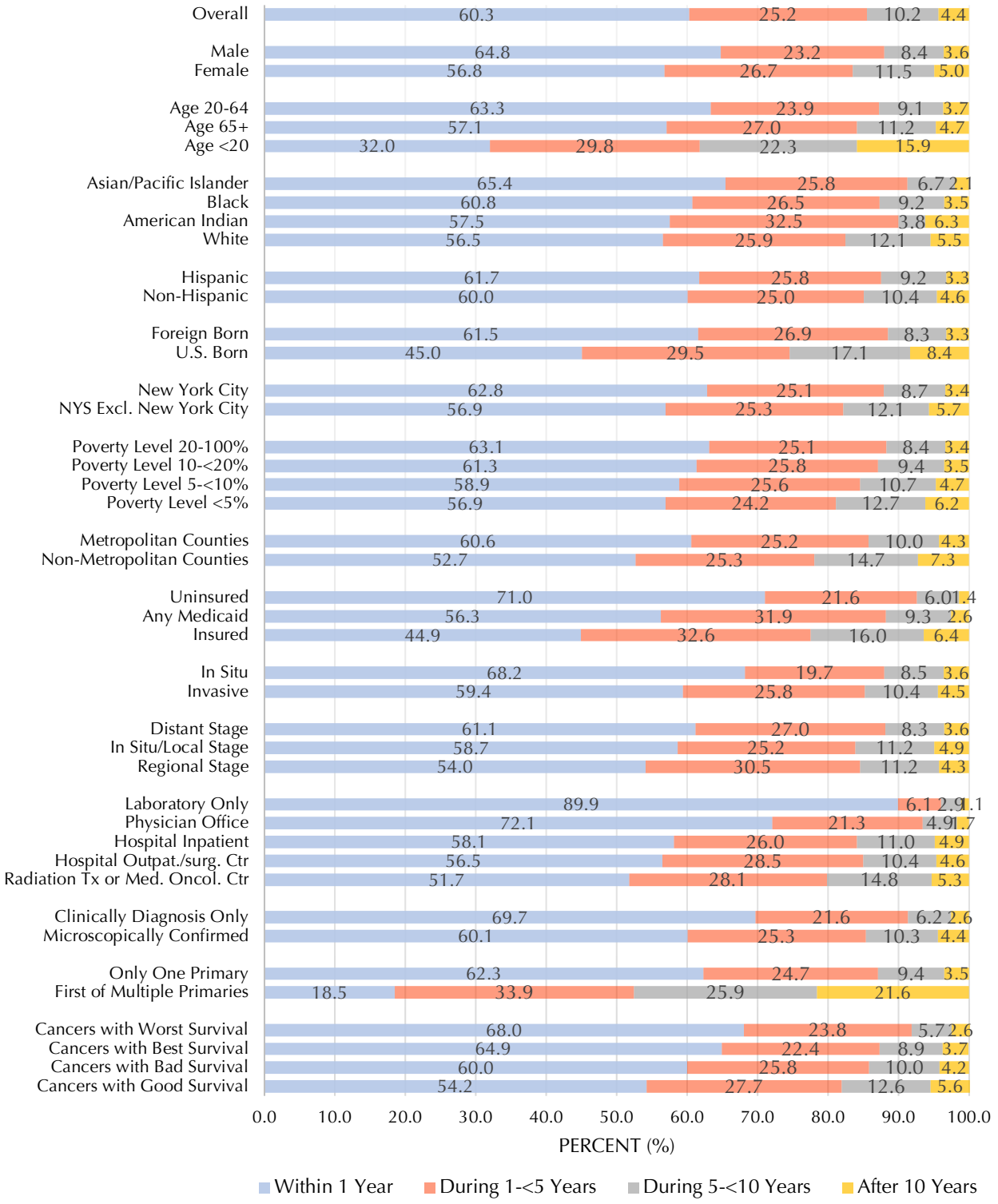
Demographic/ tumor characteristics	*Patients LTFU prior to December 31, 2018*[1]		*Patients LTFU within 5 y after cancer diagnosis*[2]	
	All, OR (95% CI)	**Alive, OR (95% CI)**	**All, OR (95% CI)**	**Alive, OR (95% CI)**
Sex				
Male	Ref	Ref	Ref	Ref
Female	1.27 (1.25–1.29)	1.22 (1.20–1.25)	1.29 (1.25–1.32)	1.24 (1.21–1.28)
Age (y)				
<20	Ref	Ref	Ref	Ref
20–64	0.43 (0.40–0.45)	0.73 (0.69-0.78)	0.61 (0.56-0.66)	1.04 (0.95-1.12)
≥65	0.35 (0.33–0.37)	0.97 (0.92-1.03)	0.34 (0.32-0.37)	0.79 (0.72-0.86)
Race				
White	Ref	Ref	Ref	Ref
Black	1.20 (1.18–1.23)	1.18 (1.16–1.21)	1.28 (1.24–1.32)	1.29 (1.25–1.33)
American Indian/Alaska Native	1.13 (0.90–1.42)	1.09 (0.87–1.38)	1.46 (1.07–1.99)	1.48 (1.08–2.02)
Asian/Pacific Islander	2.04 (1.99–2.09)	1.83 (1.77–1.88)	2.05 (1.98–2.13)	1.88 (1.81–1.95)
Unknown	3.57 (3.44–3.71)	3.31 (3.19–3.44)	4.85 (4.60–5.12)	4.56 (4.32–4.81)
Ethnicity				
Non-Hispanic	Ref	Ref	Ref	Ref
Hispanic	1.19 (1.17–1.22)	1.17 (1.14–1.20)	1.26 (1.22–1.30)	1.24 (1.20–1.29)
Birth country				
US born	Ref	Ref	Ref	Ref
Foreign born	3.61 (3.52–3.70)	2.85 (2.78–2.92)	4.97 (4.80–5.14)	4.30 (4.16–4.46)
Unknown	3.28 (3.21–3.35)	1.77 (1.73–1.80)	4.13 (4.00–4.26)	2.67 (2.59–2.75)
Region				
New York City	Ref	Ref	Ref	Ref
New York State excluding New York City	0.72 (0.70–0.73)	0.70 (0.68–0.71)	0.69 (0.67–0.71)	0.68 (0.66–0.70)
Poverty level				
0% – <5%	Ref	Ref	Ref	Ref
5% – <10%	0.93 (0.90–0.95)	0.96 (0.93–0.98)	0.93 (0.90–0.96)	0.95 (0.92–0.98)
10% – <20%	0.92 (0.90–0.94)	0.97 (0.95–0.99)	0.92 (0.89–0.95)	0.95 (0.92–0.99)
20% – 100%	0.84 (0.82–0.86)	0.93 (0.90–0.95)	0.83 (0.80–0.86)	0.89 (0.86–0.92)
Unknown	2.29 (2.06–2.55)	2.48 (2.21–2.77)	2.56 (2.25–2.92)	2.57 (2.25–2.94)
Rural–urban				
Metropolitan counties	Ref	Ref	Ref	Ref
Nonmetropolitan counties	0.74 (0.71–0.77)	0.75 (0.71–0.78)	0.64 (0.60–0.69)	0.65 (0.61–0.69)
Insurance status				
Insured	Ref	Ref	Ref	Ref
Any Medicaid	1.66 (1.63–1.70)	1.78 (1.74–1.82)	2.06 (2.00–2.13)	2.14 (2.07–2.21)
Uninsured	5.90 (5.67–6.14)	6.21 (5.96–6.48)	7.62 (7.26–8.00)	7.96 (7.56–8.38)
Unknown	2.06 (2.00–2.12)	2.19 (2.13–2.26)	3.01 (2.90–3.12)	3.35 (3.23–3.48)

Table 2, cont. Multivariate Logistic Regression Analyses Evaluating Association Between Demographic/Tumor Characteristics and Likelihood of Loss to Follow-up (LTFU), Odd Ratios, and 95% CI

Demographic/ tumor characteristics	*Patients LTFU prior to December 31, 2018[1]*		*Patients LTFU within 5 y after cancer diagnosis[2]*	
	All, OR (95% CI)	**Alive, OR (95% CI)**	**All, OR (95% CI)**	**Alive, OR (95% CI)**
Year of diagnosis				
2000–2004	Ref	Ref	Ref	Ref
2005–2009	1.25 (1.21–1.29)	0.98 (0.95–1.01)	1.13 (1.09–1.18)	1.13 (1.09–1.17)
2010–2014	1.37 (1.33–1.41)	0.86 (0.84–0.89)	1.50 (1.45–1.56)	1.46 (1.41–1.52)
2015–2018	1.57 (1.53–1.62)	0.79 (0.76–0.81)		
Tumor behavior				
In situ	Ref	Ref	Ref	Ref
Invasive	1.17 (1.13–1.20)	1.05 (1.02–1.08)	1.08 (1.04–1.13)	0.96 (0.93–1.00)
Stage				
In situ/local	Ref	Ref	Ref	Ref
Regional	0.84 (0.83–0.86)	1.00 (0.98–1.03)	0.91 (0.88–0.93)	1.10 (1.07–1.14)
Distant	0.57 (0.56–0.59)	1.14 (1.11–1.17)	0.63 (0.60–0.65)	1.48 (1.42–1.53)
Unknown	0.98 (0.96–1.01)	1.64 (1.59–1.69)	0.89 (0.86–0.93)	1.43 (1.37–1.48)
Type of reporting source				
Hospital inpatient	Ref	Ref	Ref	Ref
Radiation therapy or medical oncology center	1.19 (1.16–1.22)	0.91 (0.88–0.93)	1.06 (1.02–1.11)	0.88 (0.85–0.92)
Laboratory only	1.82 (1.74–1.90)	1.40 (1.34–1.46)	1.21 (1.14–1.28)	0.88 (0.83–0.94)
Physician office	1.92 (1.85–2.00)	1.61 (1.54–1.67)	1.18 (1.10–1.25)	0.93 (0.87–0.99)
Other hospital outpatient/ surgery center	1.17 (1.15–1.19)	0.96 (0.94–0.97)	1.21 (1.18–1.24)	1.00 (0.98–1.03)
Diagnostic confirmation				
Microscopically confirmed	Ref	Ref	Ref	Ref
Clinical diagnosis only	0.81 (0.77–0.85)	1.28 (1.21–1.35)	0.95 (0.89–1.02)	1.36 (1.27–1.47)
Unknown	0.67 (0.60–0.75)	1.08 (0.96–1.21)	0.77 (0.68–0.88)	0.96 (0.83–1.10)
Sequence number				
Only 1 primary	Ref	Ref	Ref	Ref
First of multiple primaries	0.36 (0.35–0.37)	0.31 (0.30–0.32)	0.20 (0.19–0.21)	0.15 (0.15–0.16)
Cancer site group[c]				
Cancers with best survival	Ref	Ref	Ref	Ref
Cancers with good survival	0.84 (0.82–0.86)	0.88 (0.86–0.90)	0.93 (0.90–0.96)	0.98 (0.95–1.02)
Cancers with bad survival	0.85 (0.83–0.87)	0.98 (0.96–1.00)	0.89 (0.86–0.92)	1.00 (0.96–1.03)
Cancers with worst survival	0.55 (0.53–0.56)	1.23 (1.19–1.27)	0.65 (0.62–0.68)	1.81 (1.73–1.89)

OR, odds ratio; Ref, reference. [a]Patients were diagnosed during 2000–2018 and followed through December 31, 2018. [b]Patients were diagnosed during 2000–2013 and followed for 5 years after cancer diagnosis. [c]Cancer site group was assigned according to disease survival rates. Cancers with best survival include prostate, testis, thyroid, and melanoma of the skin. Cancers with good survival include breast, cervix, uterus, bladder, and kidney. Cancers with bad survival include oral cavity, colorectal, larynx, ovary, myeloma, leukemia, and others not listed in another group. Cancers with worst survival include brain/central nervous system, esophagus, stomach, lung/bronchus, and liver/pancreatic.

Figure 1. Frequency Distributions of Timing of Loss to Follow-up by Demographic/Tumor Characteristics[a]



[a] Patients were diagnosed during 2000–2018 and followed through December 31, 2018.

aged < 20 years at cancer diagnosis, or living in New York City or metropolitan counties were more likely to be LTFU compared to the reference group. In addition, cases reported by laboratories only and physician offices had a higher likelihood of LTFU compared to those reported by hospitals. Odd ratios and 95% CIs for different demographic/tumor factors are shown in Table 2.

Overall, about 60.3% of LTFU occurred within 1 year, 25.2% from 1 to <5 years, 10.2% from 5 to <10 years, and 4.4% more than 10 years after cancer diagnosis (Figure 1). The percentage of LTFU within 1 year was particularly higher for patients who were foreign born (61.5% vs 45.0% of the US born), API (65.4% vs 56.5% of the White) or uninsured (71.0% vs 44.9% of the insured), for cases that were

reported by laboratories only (89.9%) or physician offices (72.1%) vs 58.1% of the hospital inpatient reports, and for individuals with only 1 primary (62.3% vs 18.5% for first of multiple primaries).

For the secondary analyses, among 1,304,137 patients diagnosed during 2000–2013, 37,714 were LTFU within 5 years after cancer diagnosis, representing 2.9% of all patients and 4.7% of alive patients. Overall, the findings were consistent with those observed in the main analyses, except that among alive patients, 5-year LTFU among patients < 20 years of age at diagnosis showed no statistical difference compared to those aged 20–64 years (with <20 years as a reference, odd ratio for the 20–64 years age group, 1.03; 95% CI, 0.95–1.12). In addition, cases reported only by laboratories or physician offices seemed less likely to be LTFU within 5 years after cancer diagnosis than those reported by hospitals (Table 2).

Discussion

The NYSCR has routinely conducted linkages with state vital records and the NDI to ascertain deaths for many years. Since joining the SEER program in 2018, the NYSCR has expanded its follow-up activities by linking with more data sources. Linkages with the SSA have improved the registry's follow-up rate significantly. For example, the initial SSA linkage conducted in 2019 allowed us to update follow-up information for a significant number of patients, with the overall follow-up rate jumping from 55.7% to 92.6%.[16] Furthermore, linkages with hospital discharge data and Medicaid data have helped increase the follow-up rates among patients diagnosed under age 20 years and among patients with an in situ cancer diagnosis.[17] With these additional efforts, the NYSCR has met the SEER program standards for follow-up rates; however, there were still about 4.2% of patients included in the November 2021 submission who were LTFU. As a new SEER registry, we are determined to explore new ways to improve follow-up information and survival data, and identifying which factors affect patient LTFU is an important first step.

In this study, we observed higher LTFU rates among Black, API, and Hispanic patients; these findings were consistent with previous reports.[9,10] Using the SEER 18 data, Pinheriro et al[9] have reported higher LTFU rates among Asian and Hispanic patients in comparison to non-Hispanic White patients. Multiple factors could contribute to the higher LTFU rates among those racial–ethnic groups. First, Black, API, and Hispanic patients are more likely to have a missing Social Security number (SSN) compared to Whites in cancer registry reports. Among our study population, about 11.2% of API patients, 8.6% of Hispanic patients, and 5.4% of Black patients had missing SSNs; however, only 2.7% of Whites patients had a missing SSN. Unfortunately, linkages to the SSA currently do not include individual taxpayer identification numbers, which are valid identifiers for many individuals in the United States who do not have SSNs. Second, some Asian and Hispanic individuals use different naming conventions. For example, when writing Chinese names, surnames (last name) proceed given names; for Hispanic names, compound surnames from both father and mother are often used for children. Names and SSNs are 2 critical data fields used in linkages and when they are missing, inaccurate, or not reported consistently across different data sources (ie, Chinese first name and last name swapped or different/partial Hispanic last names), linkages could be problematic, and consequently true matches might be missed. Third, some patients who are born in foreign countries might relocate to their birth country after a cancer diagnosis, and this outmigration tends to be higher among certain subgroups of Hispanic patients, probably due to cultural, economic, and political reasons.[9,10] However, currently we have no way of identifying those deaths that occur outside of the United States.

Our study showed that females were more than 20% more likely to be LTFU than males. Surname changes after marriage among women could partially explain this finding, since mismatch of last name in linkages could indicate a nonmatch, especially if the quality of other critical data items used for linkage were poor in one or both data sources.

Follow-up for childhood cancer patients has always been challenging for registries,[8] and our findings have confirmed this. In our study, about 8.8% of patients diagnosed at age < 20 years were LTFU prior to December 31, 2018, while only 5.1% of patients aged 20–64 years and 3.1% of patients aged 65 years and older were LTFU. The higher LTFU rate for children could be due to various reasons. First, data for childhood cancer patients are more likely to have missing or inaccurate SSNs. It is not uncommon that a parent's SSN is mistakenly reported in place of their child's SSN. Second, when childhood cancer patients grow into adulthood, they might move to different states for college or work. Although linkage to NDI can identify deaths that occur in other states, it is not possible to identify those still alive through the NDI. In addition, linkages with other data sources, such as hospital discharge data or claims data, do not prove fruitful after outmigration to other states. Third, some administrative data sources that are commonly used for patient follow-up might not include records for the young. The SSA data files, for example, will not include records for children who do not qualify for the program. Voter registration and Department of Motor Vehicle files, used by some SEER registries, do not provide follow-up information for childhood patients until they reach eligible age.

Our results indicate that LTFU is significantly higher among the foreign born. As discussed previously, this finding might be explained by the higher percentage of missing SSNs among foreign-born patients or a higher likelihood of migration back to their country of origin after cancer diagnosis. Additionally, some patients born in other counties might travel to New York for cancer diagnosis or treatment and be incorrectly identified as state residents. When those patients return to their resident countries, it is difficult to locate and trace them. Although the NYSCR has some quality assurance processes in place to periodically identify medical tourism, this type of error cannot be totally ruled out. We observed a much higher LTFU rate among the uninsured compared to the insured. In general, patients

without insurance coverage might be less likely to seek medical care and therefore might not be included in some data sources used to confirm patient follow-up status, such as hospital discharge or insurance claims data. Although not presented, uninsured patients in this study tended to be Black, API, Hispanic, or foreign born, and the previous explanations related to those factors could apply here, as well.

The observed higher LTFU rates for cases reported only by laboratories or physician offices could be due to data quality issues. Compared to hospital reports that are abstracted by registrars, laboratory and physician reports are more likely to have missing/unknown values for SSN, race, and ethnicity, which are commonly used for patient follow-up linkages. However, it is unclear why the likelihood of LTFU within 5 years was seemingly lower for cases reported by laboratories and physician offices in the logistic regression model based on alive patients, and we are planning further investigation.

In conclusion, identifying factors associated with patient LTFU is important for cancer registries to identify methods for improving follow-up data. This study indicates that collecting high-quality and complete demographic information is critical for conducting patient follow-up through linkages with other data sources. The study also revealed that LTFU was not random; rather, certain patient groups have higher LTFU rates than others. Although some of the factors associated with patient LTFU are not amenable to change by cancer registries, more intense follow-up of cases reported by a laboratory only or by physician offices may be a possible, although resource-intensive, solution. Linkage to LexisNexis databases, which contain public and private information on individuals, to obtain correct SSNs or residential address history could also be helpful to improve future linkages aimed to confirm follow-up status. In addition, for pediatric cancer patients, having reporting facilities resubmit their most recent follow-up information or linking with additional data sources, such as state immunization records, could potentially improve the follow-up rate for this age group. Unfortunately, there is one major limitation of this study. Procedures and extent of follow-up vary among registries and some of our findings might not be generalizable to other central registries. Additionally, due to the large size of our study population, even a small difference can be statistically significant. Finally, the 5-year LTFU rates reported in this study may be underestimated because they were based on cases with complete 5-year follow-up. However, we do not expect that the relative effects observed between patient demographic and tumor characteristics and LTFU would change substantially.

How varying LTFU rates among different patient or cancer groups affect survival estimates will be investigated in future studies, which will also examine LTFU in the context of varying approaches to survival analyses (eg, cohort vs period).

References

1. Cancer Statistics Explorer Network. Surveillance, Epidemiology, and End Results Program, National Cancer Institute. Accessed August 2022. **https://seer.cancer.gov/statistics-network**
2. US Cancer Statistics Working Group. US Cancer Statistics Data Visualizations Tool, based on 2021 submission data (1999–2019): US Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute. Released in June 2022. **https://www.cdc.gov/cancer/dataviz**
3. Johnson CJ, Wilson R, Mariotto A, et al, eds. *Cancer in North America: 2015-2019 Volume Four: Cancer Survival in the United State and Canada 2012-2018*. North American Association of Central Cancer Registries, Inc; 2022.
4. Jemal A, Ward EM, Johnson CJ, et al. Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *J Natl Cancer Inst.* 2017;109(9):djx030.
5. Allemani C, Harewood R, Johnson CJ, et al. Population-based cancer survival in the United States: data, quality control, and statistical methods. *Cancer.* 2017;123:4982-4993.
6. Wilson RJ, Ryerson AB, Zhang K, Dong X. Relative survival analysis using the Centers for Disease Control and Prevention's National Program of Cancer Registries surveillance system data, 2000–2007. *J Registry Manag.* 2014;41(2):72-76.
7. Johnson CJ, Weir HK, Fink AK, et al. The impact of National Death Index linkages on population-based cancer survival rates in the United States. *Cancer Epidemiol.* 2013;37:20-28.
8. Johnson CJ, Weir HK, Yin D, Niu X. The impact of patient follow-up on population-based survival rates. *J Registry Manag.* 2010;37(3):86-103.
9. Pinheiro PS, Morris CR, Liu L, Bungum TJ, Altekruse SF. The impact of follow-up type and missed deaths on population-based cancer survival studies for Hispanics and Asians. *J Natl Cancer Inst Monogr.* 2014;49:210-217.
10. Pinheriro PS, Callahan KE, Kobetz EN. Disaggregated Hispanic groups and cancer: importance, methodology and current knowledge. In: Ramirez AG, Trapido EJ, eds. *Advancing the Science of Cancer in Latinos [Internet].* Springer; 2020. **https://www.ncbi.nlm.nih.gov/books/NBK573253/**
11. Okuyama A, Barclay M, Chen C, Higashi T. Impact of loss-to-follow-up on cancer survival estimates for small populations: a simulation study using hospital-based cancer registries in Japan. *BMJ Open.* 2020;10:e033510.
12. Weir HK, Johnson CJ, Mariotto AB, et al. Evaluation of North American Association of Central Cancer Registries' (NAACCR) data for use in population-based cancer survival studies. *J Natl Cancer Inst Monogr.* 2014;49:198-209.
13. NPCR standards. Centers for Disease Control and Prevention National Program of Cancer Registries (NPCR) website. Reviewed August 29, 2022. Accessed September 2022. **https://www.cdc.gov/cancer/npcr/standards.htm**
14. SEER 2021 Submission Requirements and Guidelines: Attachment B. SEER Patient Follow-up Calculations (page B-1). Published September 27, 2021. Accessed August 2022. **https://seer.cancer.gov/tools/seer.nov21.instructions.pdf**
15. Brenner H, Hakulinen T. Implications of incomplete registration of deaths on long-term survival estimates from population-based cancer registries. *Int J Cancer.* 2009;125:432-437.
16. Morawski BM, Qiao B, Coyle L, Rycroft RK, Schymura MJ, Johnson CJ. Impact of linkage to the Social Security Administration on follow-up completeness and cancer relative survival estimates in 2 new SEER registries: 2000–2016 diagnosis years. *J Registry Manag.* 2020;47(2):37-47.
17. Zhang X, Kahn AR, Schymura MJ. Improvement of follow-up through linkages with state Medicaid and statewide hospital discharge data in New York. Poster presented at: North American Associations of Central Cancer Registries' 2021 Summer Forum. **https://www.naaccr.org/wp-content/uploads/2021/07/2021-combined-posters.pdf**

Original Article

The Case of the Missing 2020 Cancers: Using Claims Data to Investigate a Deficit in Incident Cancer Case Reports to the New York State Cancer Registry in 2020

Xiuling Zhang, PhD[a]; Margaret Gates Kuliszewski, ScD[a,b]; Amy R. Kahn, MS[a];Maria J. Schymura, PhD[a,b]

Abstract: Background: As the February 2022 Surveillance, Epidemiology, and End Results (SEER) Call for Data deadlines approached, the New York State Cancer Registry had received reports for approximately 10% fewer consolidated incident cases for 2020 than expected. We used claims data to examine changes in the volume of cancer claim records during the COVID-19 pandemic and possible contributors to the deficit in cancer reports. Methods: The New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) requires reporting of all patient encounters from licensed ambulatory surgery, emergency department, and hospital inpatient and outpatient providers. Each record includes patient demographics and up to 17 diagnosis codes from the *International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM). For this project, we extracted 6,725,416 SPARCS records with any malignant neoplasm code for 2018 through June 2021 for NYS residents. Using SAS 9.4, we focused on comparing the cancer-related records for 2020 to the records from 2019. Results: Overall, there were 5% more cancer-related records in 2019 than in 2018 (2,009,600 vs 1,914,364), but 8.2% fewer records in 2020 (1,844,054 total) than in 2019. Looking by month and year, the number of claims in the first 2 months of 2020 exceeded the numbers from 2019 by 5%. However, a decrease in the number of claims started in March 2020, with the biggest drop in April 2020, where there was a deficit of 38.8% for cancer-related encounter reports relative to the same month the previous year. Although the numbers rose after April, the number of claims for the last half of 2020 was still 4% lower than the same time frame in 2019. There were substantial decreases in the number of records in 2020 for all encounter types and across levels of each covariate examined, including age, sex, race/ethnicity, and facility region of NYS. In analyses of all reporting facilities, facilities in New York City had a more pronounced and more prolonged drop in reporting in 2020 than facilities in the rest of the state. Conclusions: Although SPARCS data do not provide definitive evidence of decreases in incident cancer diagnoses, these data suggest that there were fewer cancers diagnosed among NYS residents in 2020. Additional analyses are needed to assess the impacts of COVID-19-related delays in cancer diagnosis and treatment on stage at diagnosis and outcomes.

Key words: *cancer registry, cancer reporting, claims, COVID-19, New York State*

Introduction

Although it has never been easy to meet case completeness goals, the New York State (NYS) Cancer Registry (NYSCR) has always strived to attain the Centers for Disease Control and Prevention (CDC)'s National Program of Cancer Registries and, more recently, the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program's 12-month data completeness standards, the latter of which had traditionally been set at 95% based on expected number of annual incident cases. However, as the February 2022 SEER Call for Data deadlines approached, reaching 95% completeness for 2020 cases seemed impossible. Despite intense pressure on reporting facilities by field staff to maintain their regular reporting, completeness was stalled at 85% of the expected number of incident cases based on our analyses of cancer reports received and consolidated by the registry. Similarly, by the time of the submission, the registry had received approximately 10% fewer incident cancer cases for 2020 than it had received a year earlier for 2019 diagnoses. These estimates were calculated prior to provisionally finalizing the NYS incidence data, and also prior to implementation of a 2020 registry completeness estimate adjustment.[1] Potential contributors to this decrease included reporting delays at the facility level due to COVID-19-related factors or other facility-specific factors and decreases in new cancer diagnoses due to COVID-19-related delays in cancer screening and diagnosis.

The COVID-19 pandemic disrupted health care worldwide. By March 2020, the pandemic had spread globally, and areas in New York City (NYC) and NYS were heavily affected.[2-6] Between the start of the pandemic and June 1, 2020, there were 203,792 laboratory-confirmed COVID-19 cases diagnosed among residents of NYC.[5] As the number

[a]New York State Cancer Registry, New York State Department of Health, Albany, New York. [b]Department of Epidemiology and Biostatistics, University at Albany School of Public Health, Rensselaer, New York.

Address correspondence to Xiuling Zhang, New York State Cancer Registry, New York State Department of Health, 150 Broadway, Suite 361, Albany, NY 12204. Telephone: (518) 474-2255. Email: xiuling.zhang@health.ny.gov.

This work was supported in part by the Centers for Disease Control and Prevention's National Program of Cancer Registries through cooperative agreement 6NU58DP006309 awarded to the New York State Department of Health and by Contract 75N91018D00005 (Task Order 75N91018F00001) from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services. The opinions, findings, and conclusions expressed here are those of the authors and do not necessarily represent the official views of the funding agencies.

of COVID-19 cases rose throughout the country, many states and cities issued stay-at-home orders. The fear of COVID-19 infection and decreased availability of medical care due to resources being diverted to COVID-19 resulted in many people postponing their scheduled or needed medical care or not seeking medical care at all. In addition, recommendations and guidelines concerning the delay of "elective" or nonurgent procedures were issued by organizations including the CDC, the Centers for Medicare and Medicaid Service, and the American College of Surgeons.[7-10] Many states, including NYS, ordered health care facilities, physicians, and other health care providers and professionals to delay nonurgent procedures.[7] As expected, there were decreases in cancer care services during the pandemic,[11-17] including a decrease in cancer screening. We therefore sought to examine changes in cancer claims records to assess the extent to which the COVID-19 pandemic and resulting changes in health care utilization impacted the diagnosis and reporting of incident cancer cases among NYS residents for the diagnosis year 2020.

Methods

Data Source and Claims Selection

We obtained claims data for the analysis from the NYS Department of Health's Statewide Planning and Research Cooperative System (SPARCS), a comprehensive all-payer data reporting system that includes patient encounters from licensed ambulatory surgery, emergency department, and hospital inpatient and outpatient facilities. Each claims record includes patient demographics, diagnoses, treatments, services, charges, and up to 17 codes associated with each encounter from the *International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD-10-CM).[18]

We used the following criteria to select cancer-related claims to include in the analyses: NYS residents, discharge between January 1, 2018, and June 30, 2021, and ICD-10-CM diagnosis codes C00–C96 for either the principal, other, or admitting diagnosis (there are 5 diagnosis type codes on SPARCS claims records: *principal, other, external cause of injury, admitting,* and *reason for visit*). When there were multiple records for a claim, we kept only the first record per claim in the file based on the record order sequence number. A total of 6,725,416 SPARCS claims records were included in the analyses.

Since there is no direct patient identifier in SPARCS, we used a combination of unique personal identifier (UPI), date of birth, and sex to define a case in SPARCS. The UPI is a combination of the first 2 and last 2 letters of the patient's last name, the first 2 letters of the first name, and the last 4 digits of the Social Security number. When more than 1 claims record had the same UPI, date of birth, and sex, we considered them to be for the same patient. If the Social Security number component of UPI was missing, the claims were still considered to be from the same patient if they had the same first 6 characters of the UPI and the same date of birth, sex, and either patient zip code or both treating facility and medical record number.

Data Analysis

We conducted all the analyses using SAS 9.4. We examined the total number of cancer-related claims records by discharge month and year, and we calculated the percent change across time periods. We also conducted these analyses for each type of encounter (inpatient, outpatient, ambulatory surgery, or emergency department), and examined the percent change in cancer-related claims records by encounter type for 2020 compared to 2019. We used a χ^2 test to examine differences in the number of cancer-related claims for discharge year 2020 vs 2019 across encounter types.

We further examined time trends in the number of cancer-related claims records by discharge month and year for various levels of each covariate of interest. Selecting 2019 as the pre-COVID comparison year, we compared the 2020 vs 2019 records by patient age (0–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, and ≥80 years), sex (male, female, and other/unknown), race/ethnicity (non-Hispanic White, non-Hispanic Black, non-Hispanic Asian and Pacific Islander, Hispanic, and Other/Unknown), the number of encounters per patient (1, 2, 3, 4, 5–6, 7–9, 10–19, ≥20), and facility region (all facilities in NYC vs all facilities in the rest of the state based on the zip code of each facility's address). For analyses of facility region, we additionally examined changes in the number of cancer-related records for each encounter type by discharge year and changes in the total number of records by discharge month and year. We used χ^2 tests to examine the statistical significance of differences in the number of observations for each level of the covariates of interest by discharge year.

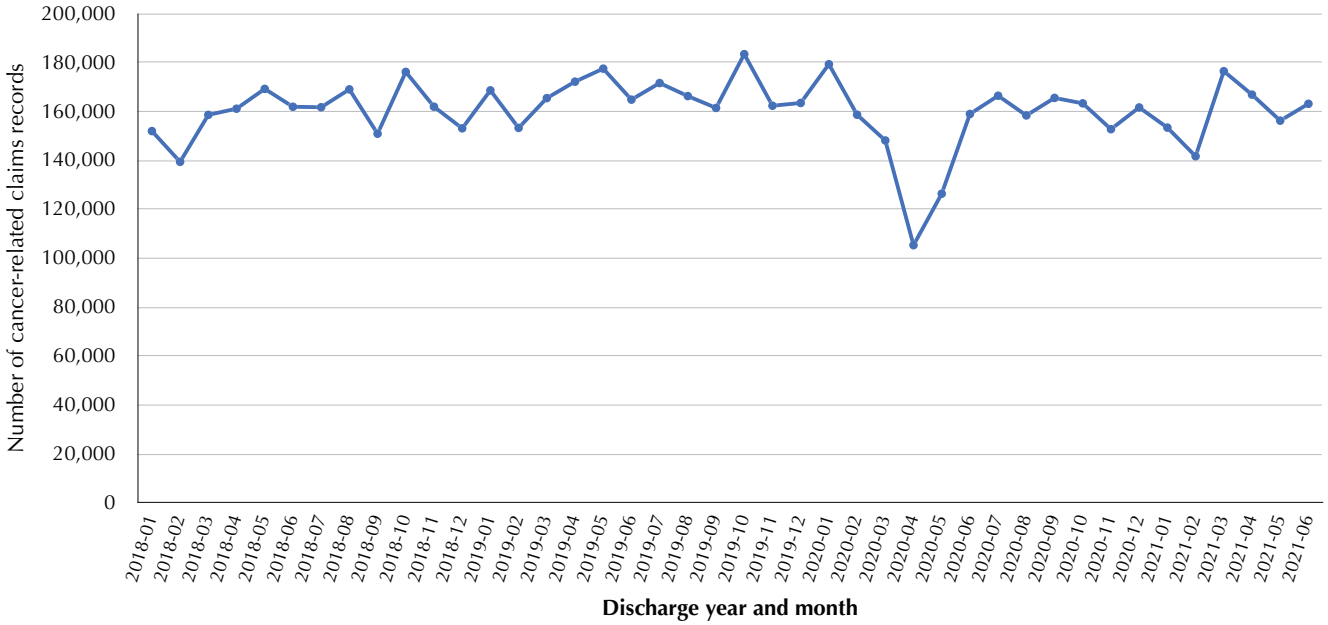
Results

Table 1 and Figure 1 show changes in the overall number of cancer-related claims over time for discharge year 2018 through June 30, 2021. In 2019, there were 5% more cancer-related records than in 2018 (2,009,600 vs

Table 1. Number of Cancer-Related Claims Records Obtained from New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) for NYS Residents by Discharge Time Frame, for Discharge Year 2018–2020

Discharge time frame	*No. of cancer-related encounters*	*% Change in no. of encounters relative to the same time frame the previous year*
2018	1,914,346	
2019	2,009,600	5.0
2020	1,844,054	–8.2
Jan–Feb 2019	321,763	
Jan–Feb 2020	337,833	5.0
April 2019	172,112	
April 2020	105,285	–38.8
Jul–Dec 2019	1,008,157	
Jul–Dec 2020	967,637	–4.0

Figure 1. Monthly Number of Cancer-Related Records Obtained from New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) for NYS Residents for Discharge Year 2018 Through June 30, 2021



1,914,346), but in 2020 the number of cancer-related records was 8.2% lower than in 2019 (1,844,054 vs 2,009,600). Looking by month and year, the number of claims in the first 2 months of 2020 exceeded the numbers from 2019 by 5%. However, a decrease in the number of claims started in March 2020, with the biggest drop in April 2020, when there was a deficit of 38.8% for cancer-related encounter reports relative to the same month the previous year. Although the numbers increased after April 2020, the number of claims for the last half of 2020 was 4% lower than the same time frame in 2019.

Table 2 and Figure 2 display changes in the number of cancer-related claims by month and year for each type of encounter. There were substantial decreases in the number of claims in 2020 compared to 2019 for all encounter types, with the biggest decrease for emergency department claims (17.4%), followed by ambulatory surgery (15.5%), inpatient claims (12.8%), and outpatient claims (6.9%). The change in the number of claims for discharge year 2020 vs 2019 was statistically significantly different across encounter types ($P < .0001$). For all encounter types, the number of claims dropped markedly in March and April 2020, increased from May 2020 until July 2020, then leveled off between July 2020 and February 2021. The number of claims remained significantly below the expected levels until March 2021 for all encounter types with the exception of inpatient claims, which had not returned to pre-COVID levels by the end of the study period. It is notable that outpatient claims returned to near pre-COVID levels fairly quickly, but the other encounter types remained lower for a longer period.

Table 2. Number of Cancer-Related Records Obtained from New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) for NYS Residents by Encounter Type for Discharge Years 2019 and 2020

Encounter type	*No. of records for 2019*	*No. of records for 2020*	*% Decrease in 2020 relative to 2019*
Ambulatory surgery	121,465	102,593	15.5
Emergency department	49,923	41,215	17.4
Inpatient	198,239	172,790	12.8
Outpatient	1,639,973	1,527,456	6.9
Total	2,009,600	1,844,054	8.2

χ^2 test: all $P < .0001$.

Table 3 displays changes in the numbers of cancer-related claims over time by age, sex, and race/ethnicity. There were substantial decreases across all levels of each covariate of interest comparing 2020 to 2019. Among 8 age groups, the largest decrease in the number of cancer-related claims was observed for individuals aged 20–29 and ≥80 years (13.1% and 11.3%, respectively), while the smallest decrease in the number of claims was observed for individuals aged 70–79 and 60–69 years (6.2% and 6.8%, respectively). These latter 2 categories accounted for over 50% of the total yearly claims in both 2019 and 2020. Looking at the number of cancer-related claims in 2020 vs 2019 among males and females, a slightly larger decrease in claims was observed in males than in females (8.6% vs 8.0% decrease). Substantial decreases in the number of claims in 2020 vs 2019 were also observed for most race/ethnicity groups, with the largest decrease for non-Hispanic Asian Pacific Islander individuals (12.7%), followed by non-Hispanic White (10.2%), and non-Hispanic Black (8.0%) individuals. Smaller decreases were observed for individuals with other or unknown race/ethnicity (4.8%) and Hispanic ethnicity (only 0.5%). There

Figure 2. Monthly Number of Cancer-Related Records Obtained From New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) for NYS Residents by Claim Type for Discharge Year 2019 Through June 30, 2021

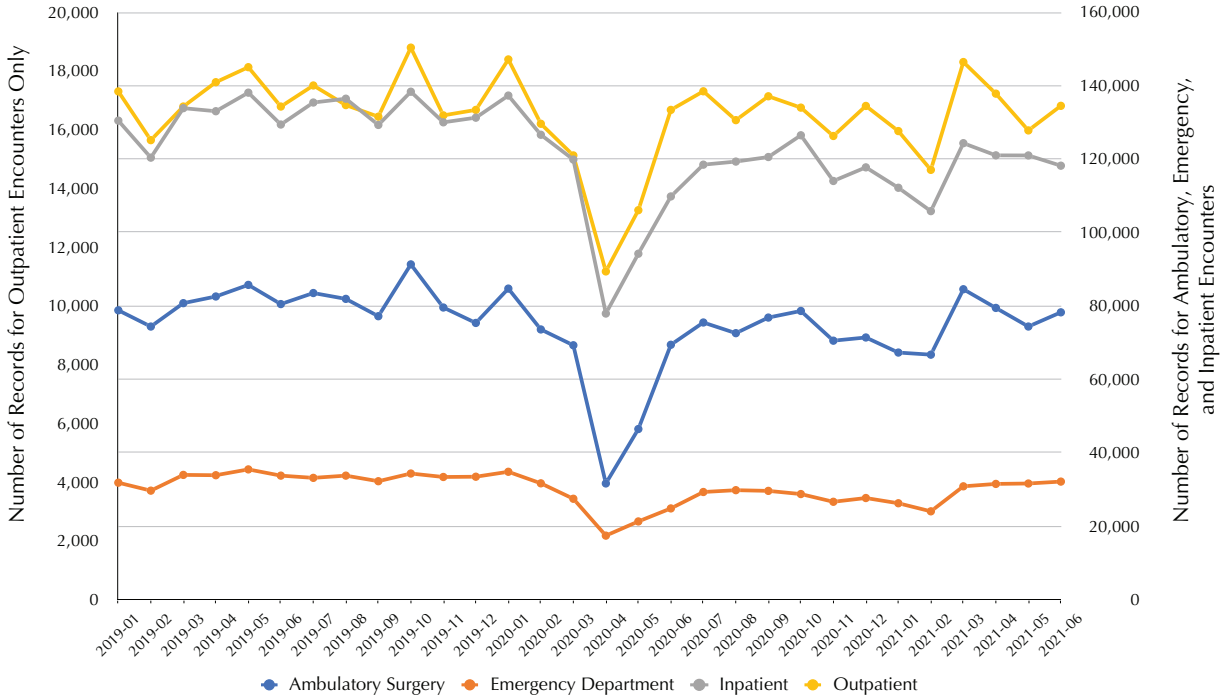


Table 3. Percent Decrease in Cancer-Related Encounters Obtained from New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) for NYS Residents for Discharge Year 2020 Compared to 2019 by Age, Sex, and Race/Ethnicity

Variable	*2019*	*2020*	*% Decrease in 2020 relative to 2019*
Age (y)			
0–19	54,808	48,934	10.7
20–29	31,649	27,506	13.1
30–39	71,966	65,277	9.3
40–49	151,828	135,020	11.1
50–59	365,424	331,787	9.2
60–69	577,358	538,353	6.8
70–79	509,789	478,212	6.2
≥80	246,778	218,965	11.3
Sex			
Female	1,051,253	967,599	8.0
Male	958,319	876,378	8.6
Other/unknown	28	77	NA
Race/ethnicity			
Non-Hispanic White	1,168,118	1,049,487	10.2
Non-Hispanic Black	258,870	238,215	8.0
Non-Hispanic API	91,728	80,119	12.7
Hispanic	206,809	205,779	0.5
Other/Unknown	284,075	270,454	4.8
Total	2,009,600	1,844,054	8.2

API, Asian/Pacific Islander; NA, not applicable. χ^2 test: $P < .0001$ for all 3 variables above.

Table 4. Number of Cancer-Related Encounters per Patient Obtained from New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) for NYS Residents for Discharge Years 2019 and 2020

No. of records per patient	*2019*	*2020*	*% Decrease in 2020 relative to 2019*
1	154,239	146,257	5.2
2	72,273	66,822	7.5
3	40,364	37,566	6.9
4	26,276	23,821	9.3
5–6	30,563	27,888	8.8
7–9	22,946	21,369	6.9
10–19	29,424	27,292	7.2
≥20	22,396	20,231	9.7
Total	398,481	371,246	6.8

χ^2 test: $P < .0001$

were statistically significant differences in the number of cancer-related encounters by discharge year across levels of all variables examined ($P < .0001$).

The number of patients with cancer-related encounters decreased by 6.8% from 2019 to 2020 (Table 4). Looking at the number of cancer-related encounters per patient, the largest decrease in the number of patients was observed for those with 20 or more claims records per year (9.7%) and the smallest decrease was observed for those with only 1 claim record per year (5.2%). The difference in the number of patients from 2019 to 2020 differed significantly by number of cancer-related encounters per patient ($P < .0001$).

As shown in Table 5 and Figure 3, facilities in NYC had a more pronounced and slightly more prolonged drop in cancer-related encounters in 2020 vs 2019 than facilities in the rest of the state. This was true for the number of claims both overall (13.5% vs 5.3% decrease in NYC vs the rest of NYS, respectively) and by type of encounter (ambulatory surgery, 26.4% vs 9.9%; emergency department, 20.9% vs 15.2%; inpatient service, 14.6% vs 11.7%; and outpatient service, 12.2% vs 3.9%).The differences in the numbers of

Table 5. Number of Cancer-Related Claims Records Obtained from New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) for NYS Residents by Encounter Type and Facility Region for Discharge Years 2019 and 2020

Claim type	*All facilities in New York City*			*All facilities in the rest of NYS*		
	Number of records		% Decrease in 2020 relative to 2019	Number of records		% Decrease in 2020 relative to 2019
	2019	2020		2019	2020	
Ambulatory surgery	41,621	30,645	26.4	79,844	71,948	9.9
Emergency department	19,591	15,506	20.9	30,332	25,709	15.2
Inpatient	78,812	67,314	14.6	119,427	105,476	11.7
Outpatient	581,674	510,573	12.2	1,058,299	1,016,883	3.9
Total	721,698	624,038	13.5	1,287,902	1,220,016	5.3

χ^2 test: P < .0001 for both all facilities in New York City and in the rest of NYS.

claims for discharge year 2020 vs 2019 were statistically significant across encounter types for facilities in NYC and for facilities in the rest of NYS ($P < .0001$).

Time trends by month and year in the numbers of cancer-related records by category of patient age, sex and race/ethnicity, number of encounters per patient, and facility region were quite similar to the overall time trend (some results not shown).

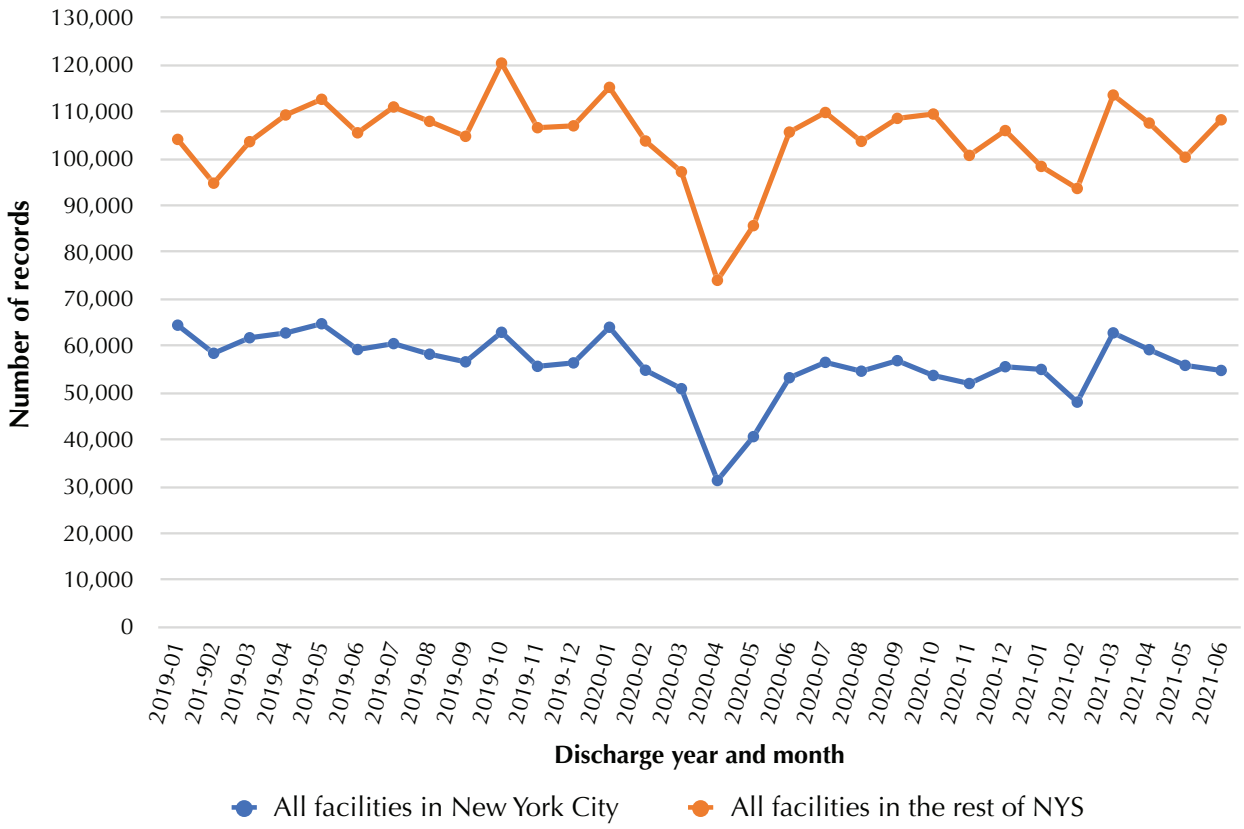
Discussion

In this analysis, we used claims data to examine changes in the volume of cancer-related services during the COVID-19 pandemic to assess the extent to which the pandemic and resulting changes in health care utilization impacted the diagnosis of incident cancer cases among NYS residents. We observed an 8.2% decrease in cancer-related claims in 2020 compared to 2019, which approximates the 10% missing cancer case reports observed leading up to the February 2022 SEER Call for Data deadlines. The additional 1.8% difference not accounted for in the statewide discharge records might in part reflect missing cases that would have been treated in nonlicensed facilities (eg, Veterans Affairs hospitals) or physician offices, neither of which are captured in SPARCS.

In time trend analyses, the decrease in the number of cancer-related claims records started in March 2020, with the biggest drop in April 2020, in which there was a deficit of 38.8% for cancer-related encounter reports relative to the same month the previous year. Although the numbers rose after April, the claims for the last half of 2020 remained 4% lower than for 2019. These changes in cancer-related claims records in 2020 compared to 2019 are consistent with the timeline of the COVID-19 pandemic, including trends in the number of daily COVID-19 cases and deaths.[19]

In analyses of changes in the number of cancer-related claims records by covariates of interest, we observed very similar time trends to the overall trends across encounter types, age groups, sex, race/ethnicity, and facility region. Analyses of the number of cancer-related encounters per patient showed decreases across all categories of the number of records per patient, and a decrease of 6.8% in the number

Figure 3. Monthly Number of Encounters Obtained from New York State (NYS) Statewide Planning and Research Cooperative System (SPARCS) for NYS Residents by Facility Region (All Facilities in New York City vs All Facilities in the Rest of NYS) for Discharge Year 2019 Through June 30, 2021



of patients with cancer-related encounters in 2020 relative to 2019. These results indicate widespread decreases in the number of cancer-related encounters as well as the number of patients seen for cancer-related encounters, supporting both a drop in the number of services and in the number of patients seeking care for cancer.

The decrease in cancer-related claim records for NYS residents in 2020 relative to 2019 is consistent with other research studies that reported drops in medical care, cancer care visits, and cancer screening rates.[11-17] Using administrative claims data, Chen and colleagues reported a 90.8% decrease in breast cancer screening, a 79.3% decrease in colorectal cancer screening, and a 63.4% decrease in prostate cancer screening for March through May 2020 compared with the same time frame in 2019. The authors estimated that these decreases corresponded to an absolute deficit of 3.9 million breast screenings, 3.8 million colorectal screenings, and 1.6 million prostate screenings across the United States during the first half of 2020.[14] A decrease in screening of this magnitude, along with delays in both routine and diagnostic medical care, would be expected to lead to a substantial deficit in the number of cancers diagnosed during the COVID-19 pandemic. Although our analysis included all types of cancer-related claims, and not just those related to a diagnosis, the analysis results still support a decrease in the number of cancer cases diagnosed in 2020. The percent decrease in the number of cancer-related claims in 2020 is consistent with the decreased number of cancer reports received by the NYSCR for diagnosis year 2020. Although it is possible that the observed decrease in cancer-related claims was due in part to existing cancer patients not receiving necessary treatment or follow-up care due to the COVID-19 pandemic, it is expected that these patients would have resumed their cancer care by the end of 2020 and would be captured in later claims data. However, our results indicate that the decrease in cancer-related claims persisted through the end of 2020.

Strengths of this study include the use of a large database of claims data with information on cancer-related claims both during and prior to the start of the COVID-19 pandemic. This allowed us to examine changes over time in the volume of cancer-related claims, as well as characteristics of the patients with cancer-related claims before and during the pandemic. However, the claims data also include data for prevalent cancer cases, and it was not possible to deduplicate the claims by person or by diagnosis. As a result, some of the change in the volume of cancer-related claims was likely related to existing cancer patients delaying cancer treatment or other care, although these patients would have been expected to resume care after the peak of the pandemic. Other limitations of the data include the unavailability of encounters from laboratories, other states, and some physicians' offices and radiation treatment centers. It therefore provides an incomplete estimate of the total number of cancer-related health encounters among New York State residents.

In summary, although SPARCS claims data do not provide definitive evidence of a decrease in incident cancer diagnoses, these data suggest that there were fewer cancers diagnosed among NYS residents in 2020. Decreases in cancer-related claims were seen across encounter types and levels of age, sex, race/ethnicity, and facility region, suggesting that there were fewer cancer diagnoses across all levels of the population and throughout NYS. Additional analyses are needed to assess the impacts of COVID-19-related delays in cancer diagnosis and treatment on stage at diagnosis and outcomes. In addition, work is needed to ensure the continuity of cancer screening, diagnostic services, and treatment during future public health emergencies to minimize delays in cancer diagnosis and care and ensure the best possible outcomes for individuals diagnosed with cancer.

References

1. Negoita S, Feuer R. *Updates to Methodology for Estimating Expected Counts of Cases Diagnosed in Year 2020 in the Context of November 2022 Data Transmission to NCI SEER: Communication to SEER Research Registry Principal Investigators and Managers*. North American Association of Central Cancer Registries; 2022. **https://share.naaccr.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=67307b9a-ca0c-4f74-8003-d770829ef7e5&forceDialog=0**
2. CDC Museum COVID-19 timeline. Centers for Disease Control and Prevention website. Accessed August 5, 2022. **https://www.cdc.gov/museum/timeline/covid19.html#:~:text=March%2011%2C%202020,declares%20COVID%2D19%20a%20pandemic**
3. Governor Cuomo announces New York ending COVID-19 state disaster emergency on June 24. New York State website. Published June 23, 2021. Accessed August 5, 2022. **https://www.governor.ny.gov/news/governor-cuomo-announces-new-york-ending-covid-19-state-disaster-emergency-june-24**
4. Bialek S, Bowen V, Chow N, et al; CDC COVID-19 Response Team. Geographic differences in COVID-19 cases, deaths, and incidence—United States, February 12–April 7, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69:465-471. doi:10.15585/mmwr.mm6915e4
5. Thompson CN, Baumgartner J, Pichardo C, et al. COVID-19 Outbreak—New York City, February 29–June 1, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(46):1725-1729. doi:10.15585/mmwr.mm6946a2
6. New York City Department of Health and Mental Hygiene (DOHMH) COVID-19 Response Team. Preliminary estimate of excess mortality during the COVID-19 outbreak—New York City, March 11–May 2, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(19):603-605. doi:10.15585/mmwr.mm6919e5
7. *Factsheet: State Action Related to Delay and Resumption of "Elective" Procedures During COVID-19 Pandemic*. American Medical Association; 2020. **https://www.ama-assn.org/system/files/2020-06/state-elective-procedure-chart.pdf**
8. *Non-Emergent, Elective Medical Services, and Treatment Recommendations*. Center for Medicare & Medicaid Service; 2020. **https://www.cms.gov/files/document/cms-non-emergent-elective-medical-recommendations.pdf**
9. *COVID-19: Recommendations for Management of Elective Surgical Procedures*. American College of Surgeons; 2020. **https://www.facs.org/media/b04pkoxp/recommendations_for_management_of_elective_surgical_procedures.pdf**
10. *ACS Guidelines for Triage and Management of Elective Cancer Surgery Cases During the Acute and Recovery Phases of Coronavirus Disease 2019 (COVID-19) Pandemic*. American College of Surgeons; 2020. **https://www.facs.org/media/xrspaqaz/acs_triage_and_management_elective_cancer_surgery_during_acute_and_recovery_phases.pdf**
11. Czeisler ME, Marynak K, Clarke KEN, et al. Delay or avoidance of medical care because of COVID-19-related concerns: United States, June 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(36):1250-1257. doi:10.15585/mmwr.mm6936a4
12. Mast C. Delayed cancer screenings—a second look. EPIC Health Research Network website. Published July 17, 2020. Accessed August 8, 2022. **https://epicresearch.org/articles/delayed-cancer-screenings-a-second-look**
13. Fedewa SA, Star J, Bandi P, et al. Changes in cancer screening in the US during the COVID-19 pandemic. *JAMA Netw Open*. 2022;5(6):e2215490. doi:10.1001/jamanetworkopen.2022.15490
14. Chen RC, Haynes K, Du S, Barron J, Katz AJ. Association of cancer screening deficit in the United States with the COVID-19 pandemic. *JAMA Oncol*. 2021;7(6):878-884. doi:10.1001/jamaoncol.2021.0884
15. Jazieh AR, Akbulut H, Curigliano G, et al; International Research Network on COVID-19 Impact on Cancer Care. Impact of the COVID-19 pandemic on cancer care: a global collaborative study. *JCO Glob Oncol*. 2020;6:1428-1438. doi:10.1200/GO.20.00351
16. Schrag D, Hershman DL, Basch E. Oncology practice during the COVID-19 pandemic. *JAMA*. 2020;323:2005-2006.
17. van de Haar J, Hoes LR, Coles CE, et al. Caring for patients with cancer in the COVID-19 era. *Nat Med*. 2020;26:665-671.
18. Statewide Planning and Research Cooperative System (SPARCS). New York State Department of Health website. Updated August 2022. **https://www.health.ny.gov/statistics/sparcs/**
19. Data timeline. John Hopkins University of Medicine Coronavirus Resource Center website. Accessed August 9, 2022. **https://coronavirus.jhu.edu/region/us/new-york**

Original Article

Determining Fitness for Use of SEER Cause-Specific Cause of Death in Analyses of Cause-Specific Survival

Bożena M. Morawski, MPH, PhD[a]; Mei-Chin Hsieh, PhD, CTR[b]; Manxia Wu, PhD[c]; Recinda Sherman, PhD, CTR[d]; Angela B. Mariotto, PhD[e]; Christopher J. Johnson, MPH[a]

Abstract: Background: Net and crude cancer survival statistics can be calculated using cause of death or expected survival from life tables. In some instances, using cause of death information may be advantageous. The Surveillance, Epidemiology, and End Results (SEER) Program cause-specific cause of death variable (North American Association of Central Cancer Registries [NAACCR] item #1914) designates that a patient died of their cancer. We evaluated how missingness in NAACCR item #1914 impacted survival estimates to determine fitness for use in NAACCR Cancer in North America (CiNA) products. Methods: We used CiNA survival and prevalence data (November 2020 submission) to calculate 60-month cause-specific survival among persons aged 15–99 years at time of diagnosis using NAACCR item #1914. We treated missing/unknown causes of death in 3 ways: excluded from analysis, included as dead from this cancer, or included as censored at time of last follow-up. Autopsy/death-certificate-only cases were excluded from survival analyses. We calculated the proportion of deaths with unknown/missing cause of death by registry and demographic variables. Results: Generally, 60-month cause-specific survival estimates differed by <1% between the 3 approaches when NAACCR item #1914 was missing/unknown for <3% of deaths. When applying a <3% fit-for-use standard to SEER cause-specific cause of death, data from 34 registries were included in cause-specific survival analyses. The proportion of deaths with missing/unknown cause of death varied by primary site, age at diagnosis, race/ethnicity, year of diagnosis, and registry. Conclusions: We have identified missingness cut points for NAACCR item #1914, which strike a balance between scientific integrity and registry inclusiveness, to designate data in NAACCR CiNA data products as fit for use in cause-specific survival analyses.

Key words: *survival estimates, cause-specific cause of death variable, cause of death*

Background

Each year, members of the North American Association of Central Cancer Registries (NAACCR) voluntarily submit data to develop an aggregated resource for cancer surveillance and research. This aggregated resource is used to create multiple data products, including the Cancer in North America (CiNA) monographs,[1-5] and the CiNA research data set. For inclusion in incidence data products, registry data must meet certification criteria for submission timeliness, completeness, and accuracy. The highest level of certification is given to registries that meet the following criteria: case ascertainment of ≥95% completeness; <3% of cases are only identified via a death certificate; <0.1% of tumors are duplicates per the NAACCR duplicate protocol; all fields used to calculate incidence statistics (cancer type, sex, race, age, and county) are error-free (ie, pass edits); <2% of tumors are missing meaningful information on age, sex, and county; <3% of tumors are missing meaningful information on race (United States only); and the file is submitted to NAACCR within 23 months of the end of the submission diagnosis year.[6]

For inclusion in the CiNA survival and prevalence volumes and data sets, registries must additionally meet Surveillance, Epidemiology, and End Results (SEER) Program standards for follow-up or ascertain all deaths through the study cutoff date (including state mortality file linkage and National Death Index [NDI] linkage for US registries and provincial/territorial mortality file linkage for Canadian registries). In 2021, these criteria—which are applied to the years of data included in survival and prevalence estimates—received their own NAACCR recognition, "Fitness for Use for Survival & Prevalence Recognition." The criteria for this recognition only pertain to overall vital status and not cause of death. Accordingly, current recognition criteria are well aligned with the most commonly used population-based cancer survival statistics that do not require cause of death information: relative survival ratios and the Pohar-Perme estimator.[7]

Net cancer survival can be calculated using a relative or cause-specific survival approach. Relative survival estimates represent the ratio of the observed-to-expected survival among cancer patients in the absence of competing causes of death, where expected survival is determined from matched life tables. Relative survival methods have the advantage of not requiring cause of death information from death certificates, which can be inaccurate.[8,9] Additionally, relative survival has the advantage of representing any excess mortality experienced by cancer patients (eg, late

[a]Cancer Data Registry of Idaho, Idaho Hospital Association, Boise, Idaho. [b]Louisiana Tumor Registry and Epidemiology Program, Louisiana State University Health Sciences Center School of Public Health, New Orleans, Louisiana. [c]Centers for Disease Control and Prevention, Atlanta, Georgia. [d]North American Association of Central Cancer Registries, Springfield, Illinois. [e]National Cancer Institute, Bethesda, Maryland.

Address correspondence to Bożena M. Morawski, MPH, PhD. Email: bmorawski@teamiha.org.

cardiotoxic effects of therapy as opposed to mortality attributable to only the cancer). The accuracy of relative survival measures can, however, be greatly influenced by the appropriateness of life tables selected to represent the expected survival of the study population.[10-12] Life tables should be matched on factors influencing cancer-specific and overall survival, such as age, sex, race/ethnicity, socioeconomic status, geography,[10] and factors relevant to the cancer under study.[13] In instances where life tables are not well matched to the study population, cause-specific survival may be more informative to researchers, public health professionals, and policymakers than relative survival. Appropriate life tables may be difficult to find for study populations that are not well characterized (eg, calculating survival for screening-related cancers that may be diagnosed among populations that are healthier than the general population, or for cancers where incidence and mortality are related to underlying risk factors such as smoking). Finally, large shifts in population-level mortality due to specific events (eg, decreases in life expectancy at birth due to COVID-19 or the US opioid crisis) may render previously appropriate life tables inappropriate.

As there is increased interest in estimating cause-specific survival at the population level,[8,14,15] it becomes increasingly important to ensure that researchers have access to high-quality, population-level cause of death information. More precisely, it is important that researchers be aware of jurisdiction-specific limitations of population-level data—in particular, when they intend to present regional, registry-specific, or other subpopulation estimates. For example, some registries may not be able to release cause of death information to national bodies (ie, NAACCR, SEER, and the National Program of Cancer Registries [NPCR]), resulting in large-proportions of jurisdiction-specific missingness. Registry operations—in particular, timing and sources of linkages—also impact the completeness of cause of death information, whereby a particular follow-up or diagnosis year is missing a higher-than-average proportion of cause of death among deceased cancer patients. There are numerous approaches to handling missing data, including cause of death, and some of these ad hoc methods have been shown to bias estimates.[14,16-18] Acknowledging, however, that many researchers conduct analyses of cancer surveillance data (eg, CiNA, SEER) in SEER*Stat alone necessitates the identification of data that are fit for use in cause-specific survival upstream of data release to researchers. The impact of cause of death missingness on survival estimates as calculated in SEER*Stat has not been evaluated in the NAACCR CiNA data set.

This study used the CiNA survival and prevalence data to assess how the proportion of missingness in the SEER cause-specific cause of death variable (NAACCR item #1914) impacted survival estimates and establishes fit-for-use cut points indicating when data should be excluded from cause-specific survival analyses. The ultimate goal of these efforts is to increase visibility around completeness of cause of death information and improve the completeness of cause of death data, such that no otherwise qualifying tumors will be censored or excluded from analyses due to missing cause of death information.

Methods

We evaluated the impact of excluding registries with >0.5%, >2%, and >3% missing/unknown SEER cause-specific cause of death on (1) the number of registries that would be included in cause-specific survival analyses and (2) the survival estimates themselves. We used NAACCR CiNA survival and prevalence data for the United States and Canada (December 2020 submission)[19] to calculate 60-month cause-specific survival among persons aged 15–99 years at time of diagnosis with a malignant tumor during 2011–2017. Following the methods in the NAACCR CiNA monograph, estimates were age-standardized using the International Cancer Survival Standards, which include patients diagnosed at ages 15–99 years.[4] Follow-up time was calculated using a blended approach.[4] For registries meeting SEER follow-up standards, follow-up time was calculated through the first of date of death, date of last contact, or end of the study period (December 31, 2017). For all other registries, the presumed alive method was used, meaning follow-up time for patients not known to be deceased was calculated through the end of the study period. For registries conducting active follow-up (ie, ascertaining vital status and date of last contact via linkages with administrative or hospital databases), alive cases with no follow-up time were excluded (about 0.17%). Tumors reported only via death certificate or autopsy were excluded, but proportions of these cases were evaluated by primary site, patient demographics, and registry. Cause of death was determined by the SEER cause-specific cause of death field (NAACCR item #1914),[20] which is described as follows: To capture deaths related to the specific cancer but not coded as such, the SEER cause-specific death classification variables are defined by taking into account causes of deaths in conjunction with tumor sequence (ie, only 1 tumor or the first of subsequent tumors), site of the original cancer diagnosis, and comorbidities (eg, AIDS and/or site-related diseases).[13,15] Other survival analysis parameters matched those used in the CiNA survival volume (**https://www.naaccr.org/wp-content/uploads/2022/06/CiNA.2015-2019.v4.survival.pdf**).[4]

Survival calculations were performed on blended survival time in SEER*Stat version 8.4.0.1 using the actuarial method.[21] Three sets of survival estimates were calculated by classifying tumors with missing/unknown cause of death in the following ways: (1) excluding these tumors from analyses; (2) including these tumors in analyses with a cause of death of the cancer under study; (3) including these tumors in analyses and censoring them at time of death, with the assumption that the cause of death is not cancer. All eligible tumors were included in analyses vs restricting to first primary.[15] We calculated proportion of deaths with unknown or missing cause of death codes by registry, primary site, patient demographics (age, race/ethnicity, rural vs urban residence at time of diagnosis), and other covariates. We then compared absolute differences in 60-month cause-specific survival estimates as calculated by each method described in (1)–(3) above, with particular attention to instances where missingness in cause of death yielded ≥1% absolute difference in survival estimates.

Among registries with <3% missing/unknown SEER cause-specific cause of death, we quantified differences in the proportion missing/unknown SEER cause-specific cause of death by registry and age at diagnosis (15–64 years vs ≥65 years), race/ethnicity (race and origin recode: non-Hispanic white, non-Hispanic Black, non-Hispanic American Indian/Alaska Native, non-Hispanic Asian or Pacific Islander, Hispanic), county-level urban/rural residence at diagnosis (2013 Beale codes), type of reporting source, follow-up source central, and primary site (SEER site recode ICD-O-3/WHO 2008) to identify potential factors driving missing/unknown cause of death. Follow-up source central (NAACCR item #1791) indicates the source of consolidated vital status, date of last contact, and cause of death information, as applicable. Finally, we examined patterns of missing/unknown cause of death by year of diagnosis within registry to describe how the proportion of deaths with missing/unknown cause may vary by changes in registry practice or policy over time.

Results

We evaluated data from 58 central cancer registries (50 US states, the District of Columbia, 7 provincial Canadian registries) receiving the CiNA "Fitness for Use for Survival & Prevalence Recognition." The percent missing/unknown SEER cause-specific cause of death among 11,757,022 tumors diagnosed during 2012–2017 with follow-up through the end of 2017 ranged by registry, from 100% in 6 US or Canadian registries to 0.02% for 1 US registry; the median proportion of missing/unknown SEER cause-specific cause of death across registries was 1.52% (Table 1). We saw evidence that missingness in cause of death information was impacted by year-to-year variation in registry operations. Within registry, missingness varied across diagnosis years (data not shown). Among registries where cause of death was missing for <100% and >3% of tumors for all study years combined, the number of years with <3% missing/unknown SEER cause-specific cause of death information ranged from 0 to 6 of the 7 diagnosis years. There was a high degree of correlation between which registries had ≥3% missing/unknown SEER cause-specific cause of death overall and those registries with >10% missing/unknown SEER cause-specific cause of death for a specific year, indicating that >10% missing/unknown SEER cause-specific cause of death for a specific year was an additional informative marker of biased survival estimates (data not shown).

As the percentage of tumors with missing/unknown SEER cause-specific cause of death information increased, absolute differences in all sites combined survival estimates using the 3 methods also increased (Table 1, Figure 1). For 10 registries with <100% and >3% missing/unknown SEER cause-specific cause of death, the median absolute difference in cause-specific survival estimates was 0.2% (interquartile range [IQR], 0.1–1.0) between methods that censored patients with unknown cause of death at date of last follow0-up vs excluded them, 1.5% (IQR, 1.2–11.5) between methods that censored patients with unknown cause of death at date of last follow-up vs included these patients as dead from their cancer, and 1.3% (IQR, 1.1–10.5) between methods that censored patients with unknown cause of death at date of last follow-up vs excluded them, 1.4% (IQR, 0.9–14.5) between methods that censored patients with unknown cause of death at date of last follow-up vs included these patients as dead from their cancer, and 1.2% (IQR, 0.9–13.3) between methods that excluded patients from analyses vs classifying these patients as dead from their cancer. An inclusion cut point of <3% missing/unknown SEER cause-specific cause of death strikes a balance between scientific integrity in survival estimates and registry inclusiveness; ie, a minimal number of registries are excluded from 5-year cause-specific survival calculations and differences in estimates from included registries demonstrated ≤1% differences in 5-year cause-specific survival by method as calculated in SEER*Stat. The selection of other cutoff points (eg, missingness of <2% or <0.5%) yielded smaller median differences in survival estimates (Table 1). However, these cut points exclude 22 and 9 additional registries, respectively, from cause-specific survival analyses, with minimal corresponding benefit in reducing bias in cause-specific survival estimates. Thus, registries meeting the <3% missing/unknown cut point were deemed fit for use.

Subsequent analyses evaluated patterns of missingness using data from 34 US registries deemed fit for use (<3% missing/unknown SEER cause-specific cause of death). Substate registries covered by their entire state were excluded from these analyses (ie, Greater California, Greater Bay, Los Angeles, and Seattle) and substate registries not covered by their entire state were included (ie, Detroit). An examination of the percent of tumors missing cause of death information by primary site (Table 2) demonstrated that cancers of the blood (ie, leukemia, Hodgkin lymphoma, and myeloma) had the highest mean proportions of missing cause of death (3.7%, 2.7%, and 2.1%, respectively). Tumors of the larynx (9.1%), liver and bile duct (5.4%), and stomach (5.1%) also had particularly high maximum values of missingness, indicating that identifying cause of death for these cancers might be more difficult in some registry jurisdictions. Cause-specific survival estimates for specific primary sites were impacted less when not stratifying by registry, with the largest differences in survival estimates being 0.2 for liver and intrahepatic bile duct and stomach, 0.6 for stomach, and 0.4 for stomach and cervix uteri between methods.

Given the differences in cause of death completeness by primary site and registry, we further analyzed tumor missingness by patient demographic characteristics and registry among the 34 registries meeting the <3% overall cause of death missingness criterion (Table 3). The mean and median values of cause of death missingness by age category (age 15–64 years at diagnosis vs age ≥65 years at diagnosis) were similar (median, 1.04% in 15–64 years vs 0.85% in ≥65 years; mean, 1.22% in 15–64 years vs 1.00% in ≥65 years). Proportion of missingness was also similar across diagnosis year and urbanicity of county of residence at time of diagnosis. Differences in missingness by race/ethnicity were larger, with the median proportion of missingness of 3.46% among Hispanic patients (all races) and

Table 1. Ranking of Percent Missing Cause of Death and Impact on Cause-Specific 5-Year Survival Estimates, 2011–2017 Diagnosis Years (Site Recode ICD-O-3/WHO 2008, All Sites Combined)

NAACCR registry number	*% DCO/ autopsy*	*% Missing COD*	*n*	*5-year cause-specific survival*			*Absolute difference in survival estimates*		
				Censored	**Excluded**	**Cancer death**	**Censored–excluded**	**Censored–dead**	**Excluded–dead**
Excluded per >3% missing/unknown criterion (registry n = 16)									
60	0.74	–	164,106	100.0	100.0	100.0	0.0	0.0	0.0
62	1.22	–	5,988	100.0	100.0	100.0	0.0	0.0	0.0
11	1.82	100.0	436,323	100.0	100.0	58.2	0.0	41.8	41.8
55	0.34	100.0	31,972	100.0	100.0	100.0	0.0	0.0	0.0
59	0.71	100.0	23,203	100.0	100.0	100.0	0.0	0.0	0.0
58	0.81	100.0	36,351	100.0	100.0	100.0	0.0	0.0	0.0
54	0.28	86.03	43,200	93.1	90.8	54.7	2.3	38.4	36.0
39	0.72	81.48	17,594	91.5	89.6	59.7	1.9	31.8	29.9
40	2.29	48.75	39,187	79.6	78.5	65.2	1.2	14.5	13.3
50	3.36	9.57	195,327	71.0	70.8	68.6	0.3	2.5	2.2
44	0.18	8.00	98,293	70.3	70.1	68.8	0.2	1.6	1.4
19	1.41	6.85	351,005	66.8	66.8	66.6	0.0	0.2	0.2
35	2.92	5.25	136,701	68.8	68.7	67.6	0.1	1.2	1.1
37	1.98	4.73	24,074	71.9	71.9	71.0	0.1	0.9	0.9
23	1.87	4.49	170,860	64.4	64.2	63.1	0.2	1.3	1.1
38	1.60	4.31	328,945	71.0	70.7	69.6	0.2	1.4	1.2
Excluded per >2% missing/unknown criterion (registry n = 22)									
16	0.83	2.56	29,752	66.8	66.8	66.7	0.0	0.0	0.0
271	1.02	2.50	211,430	70.8	70.7	70.1	0.1	0.6	0.5
4	3.05	2.42	197,943	65.7	65.5	64.9	0.1	0.7	0.6
272	1.01	2.30	257,052	67.1	67.0	66.4	0.1	0.7	0.6
3	0.92	2.18	134,915	70.2	70.1	69.5	0.1	0.6	0.5
7	1.03	2.16	24,374	69.9	69.9	69.4	0.1	0.6	0.5
27	1.19	2.11	1,077,506	68.2	68.1	67.6	0.1	0.6	0.5
41	2.30	1.94	147,373	70.8	70.7	70.4	0.1	0.4	0.4
273	1.32	1.90	608,521	67.6	67.6	67.1	0.1	0.6	0.5
21	2.58	1.77	702,655	67.2	67.1	66.7	0.1	0.5	0.4
12	0.95	1.72	173,361	62.7	62.7	62.4	0.1	0.3	0.2
13	1.53	1.71	18,573	67.3	67.2	66.8	0.1	0.5	0.4
1	2.75	1.57	105,279	64.1	64.0	63.6	0.1	0.5	0.4
42	2.67	1.53	59,198	64.5	64.4	64.0	0.1	0.5	0.4
56	0.19	1.51	117,475	67.8	67.7	67.3	0.1	0.5	0.4
20	1.75	1.49	208,898	68.3	68.3	68.0	0.0	0.3	0.3
15	1.43	1.49	63,098	67.7	67.6	67.3	0.1	0.4	0.3
5	1.32	1.44	37,221	67.3	67.2	67.0	0.1	0.3	0.3
26	1.07	1.43	720,505	69.7	69.7	69.3	0.1	0.4	0.3
33	1.40	1.37	46,069	67.4	67.3	67.0	0.1	0.4	0.3
311	0.57	1.09	165,574	70.4	70.3	70.1	0.0	0.3	0.3
18	1.19	1.05	161,946	64.8	64.7	64.4	0.0	0.3	0.3

Table 1, cont. Ranking of Percent Missing Cause of Death and Impact on Cause-Specific 5-Year Survival Estimates, 2011–2017 Diagnosis Years (Site Recode ICD-O-3/WHO 2008, All Sites Combined)

NAACCR registry number	*% DCO/ autopsy*	*% Missing COD*	*n*	*5-year cause-specific survival*			*Absolute difference in survival estimates*		
				Censored	**Excluded**	**Cancer death**	**Censored–excluded**	**Censored–dead**	**Excluded–dead**
Excluded per >0.5% missing/unknown criterion (registry n = 9)									
9	2.43	0.95	506,900	67.3	67.3	67.1	0.0	0.2	0.2
17	3.37	0.93	125,652	62.9	62.9	62.7	0.0	0.2	0.2
34	2.18	0.86	55,787	66.2	66.2	66.0	0.0	0.2	0.2
6	1.41	0.86	315,450	66.9	66.8	66.6	0.0	0.2	0.2
461	1.51	0.85	146,412	67.1	67.1	66.9	0.0	0.3	0.2
49	3.64	0.83	414,100	65.9	65.9	65.7	0.0	0.2	0.2
36	1.96	0.74	186,434	70.4	70.4	70.2	0.0	0.2	0.2
29	2.51	0.62	232,863	64.5	64.5	64.3	0.0	0.2	0.1
25	0.57	0.62	66,474	73.1	73.0	72.9	0.0	0.2	0.1
Never excluded (registry n = 11)									
24	1.72	0.50	52,815	70.5	70.5	70.4	0.0	0.1	0.1
31	1.61	0.50	231,496	69.7	69.7	69.6	0.0	0.1	0.1
47	2.44	0.45	212,591	64.0	64.0	63.9	0.0	0.1	0.1
8	2.09	0.42	51,483	67.3	67.3	67.2	0.0	0.1	0.1
28	1.97	0.41	38,100	66.5	66.4	66.3	0.0	0.1	0.1
10	0.99	0.33	104,837	62.4	62.3	62.3	0.0	0.1	0.1
2	1.67	0.28	114,955	66.2	66.2	66.2	0.0	0.1	0.1
32	0.89	0.23	77,156	62.2	62.2	62.1	0.0	0.1	0.0
53	1.53	0.22	496,534	67.5	67.5	67.5	0.0	0.1	0.0
45	2.39	0.21	172,339	65.1	65.1	65.0	0.0	0.0	0.0
30	1.61	0.02	783,858	67.5	67.5	67.5	0.0	0.0	0.0

COD, cause of death; DCO, death certificate only; ICD-O-3, International Classification of Diseases for Oncology, Third Edition; NAACCR, North American Association of Central Cancer Registries; WHO, World Health Organization. NAACCR member registries listed above include state, metropolitan, provincial, and territorial registries.

Table 2. Percent of Survival Records Missing Cause of Death (Maximum and Mean) and Survival Statistics for 34 US Registries with < 3% Missing/Unknown Cause-Specific Cause of Death by Primary Site for Diagnosis Years 2011–2017

Primary site at diagnosis	*Max*	*Mean*	*Survival, n*	*Cause spec censored*	*Cause spec exclude*	*Cause spec dead*	*Diff censored exclude*	*Diff censored dead*	*Diff exclude dead*
All sites	2.4	1.1	7,048,111	67.0	66.9	66.7	0.0	0.3	0.3
Oral cavity and pharynx	2.8	0.8	199,700	67.9	67.9	67.6	0.0	0.3	0.3
Esophagus	4.5	1.7	77,217	23.3	23.2	23.1	0.1	0.2	0.2
Stomach	5.1	1.5	106,817	35.6	35.4	35.1	0.2	0.6	0.4
Colon and rectum	2.8	1.2	632,661	64.7	64.7	64.4	0.1	0.3	0.3
Liver and intrahepatic bile duct	5.4	1.5	145,021	23.4	23.2	22.9	0.2	0.5	0.3
Pancreas	1.8	0.6	209,139	12.1	12.0	11.9	0.1	0.2	0.1
Larynx	9.1	1.7	57,895	65.7	65.7	65.4	0.1	0.3	0.3
Lung and bronchus	3.3	1.2	959,490	26.2	26.1	25.9	0.1	0.3	0.2
Melanoma of the skin	4.6	1.2	351,506	89.6	89.6	89.5	0.0	0.1	0.1
Breast	3.5	1.2	1,078,023	88.6	88.5	88.4	0.0	0.2	0.2
Cervix uteri	2.9	1.2	60,254	68.6	68.5	68.1	0.1	0.4	0.4
Corpus and uterus, NOS	1.9	0.7	240,459	81.0	81.0	80.8	0.0	0.2	0.2
Ovary	4.6	1.7	96,486	49.4	49.4	49.1	0.1	0.3	0.2
Prostate	2.9	0.5	895,181	92.7	92.7	92.5	0.0	0.2	0.2
Testis	3.8	1.2	40,157	95.3	95.3	95.3	0.0	0.1	0.1
Urinary bladder	2.8	1.2	329,458	77.2	77.2	76.9	0.0	0.3	0.2
Kidney and renal pelvis	2.8	1.2	275,887	77.5	77.5	77.2	0.0	0.3	0.2
Brain and other nervous system	4.3	0.9	92,316	27.7	27.6	27.3	0.1	0.3	0.2
Thyroid	1.8	0.7	213,979	97.1	97.1	97.0	0.0	0.1	0.1
Hodgkin lymphoma	5.6	2.7	38,400	87.2	87.2	87.0	0.0	0.2	0.2
Non-Hodgkin lymphoma	3.0	1.0	315,332	72.5	72.5	72.2	0.0	0.3	0.3
Myeloma	12.5	2.1	114,643	59.7	59.6	59.3	0.1	0.3	0.3
Leukemia	12.8	3.7	219,241	61.1	61.0	60.7	0.1	0.3	0.3
Mesothelioma	3.4	1.4	14,752	12.1	12.0	11.9	0.1	0.2	0.1

Max, maximum; NOS, not otherwise specified; spec, specific.

Figure 1. Scatterplot of Percent Missingness in Cause of Death vs Absolute Difference in Cause-Specific Survival Estimates (Censored vs Dead from This Cancer)

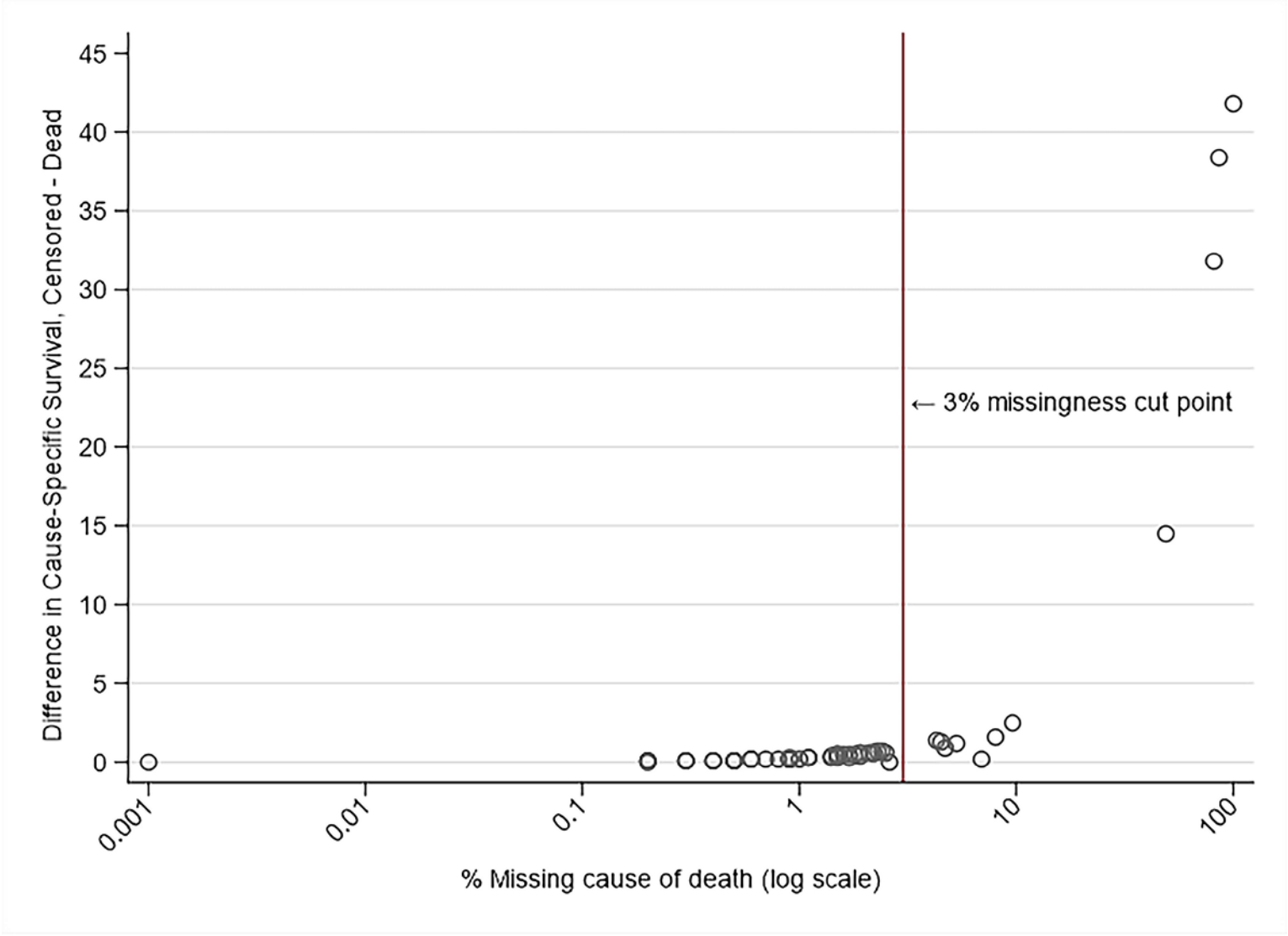


Table 3. Maximum, Mean, and Median Percent Missing COD by Demographic Characteristics, All Sites Combined, for 34 US Registries with < 3% Missing SEER Cause-Specific Cause of Death (Site Recode ICD-O-3/WHO 2008, All Sites) for Diagnosis Years 2011–2017

Demographic category	*Diagnosis year*	*Maximum (%)*	*Mean (%)*	*Median (%)*
Age group (y)				
15–64	2011–2017	2.97	1.22	1.04
≥65	2011–2017	2.28	1.00	0.85
Race and ethnicity				
Hispanic (all races)	2011–2017	5.22	3.13	3.46
Non-Hispanic American Indian/Alaska Native	2011–2017	6.67	1.28	0.90
Non-Hispanic Asian or Pacific Islander	2011–2017	23.08	4.55	3.79
Non-Hispanic Black	2011–2017	15.38	1.90	1.25
Non-Hispanic White	2011–2017	2.06	0.85	0.79
Urban/rural (2013)				
Metropolitan counties	2011–2017	2.93	1.10	0.87
Counties in metropolitan areas ≥ 1 million population	2011–2017	4.92	1.23	0.95
Counties in metropolitan areas of 250,000 to 1 million population	2011–2017	2.26	1.02	0.90
Counties in metropolitan areas of < 250,000 population	2011–2017	2.93	1.09	0.94
Nonmetropolitan counties	2011–2017	2.62	1.02	0.88
Urban population of ≥ 20,000 adjacent to a metropolitan area	2011–2017	2.71	0.87	0.62
Urban population of ≥ 20,000 not adjacent to a metropolitan area	2011–2017	6.39	1.24	0.74
Urban population of 2,500 to 19,999, adjacent to a metropolitan area	2011–2017	2.98	1.00	0.91
Urban population of 2,500 to 19,999, not adjacent to a metropolitan area	2011–2017	3.15	1.06	0.89
Comp rural < 2,500 urban population, adjacent to a metropolitan area	2011–2017	6.25	1.17	0.81
Comp rural < 2,500 urban population, not adjacent to metropolitan area	2011–2017	3.31	1.01	0.90
Missing or unknown state/county includes XX, YY, ZZ or 999	2011–2017	33.33	7.50	3.35
By year				
	2011	3.23	1.16	1.06
	2012	3.07	1.07	0.99
	2013	3.37	1.03	0.94
	2014	2.39	1.00	0.86
	2015	4.36	1.11	0.85
	2016	2.85	1.00	0.79
	2017	3.01	1.18	0.98

COD, cause of death; ICD-O-3, International Classification of Diseases for Oncology, Third Edition; SEER, Surveillance, Epidemiology, and End Results Program; WHO, World Health Organization.

3.79% in non-Hispanic Asian or Pacific Islander patients vs 0.90% in non-Hispanic American Indian/Alaska Native patients, 1.25% in non-Hispanic Black patients, and 0.79% in non-Hispanic White patients. We noted that the proportion missing/unknown also differentially impacted survival estimates by method within registry by race/ethnicity, with the largest median differences in estimates using dead from their cancer vs censored as the cause of death—a difference of 1.1%—among non-Hispanic Asian or Pacific Islander patients and Hispanic patients (all races) (Table 4).

Discussion

This is the first time that the impact of registry-specific cause of death missingness on survival estimates has been evaluated using data for the United States and Canada. This study used CiNA survival data to establish fit-for-use criteria that indicate when data from specific registries should be excluded from CiNA data products that intend to present cause-specific survival. The results of these analyses, which were based on the November 2020 data submission, support a recommendation that registries should be deemed fit for use for cause-specific survival analyses when <3% of tumors have missing/unknown SEER cause-specific cause of death.

Additionally, we noted how patterns in cause of death ascertainment mirror issues in ascertainment of vital status overall,[22,23] illustrating how important it is for researchers to investigate patterns of missingness in SEER cause-specific cause of death for their specific study questions. Differential cause of death missingness was noted in specific racial/ethnic populations and for specific primary sites. For example, a higher proportion of SEER cause-specific cause of death was missing for persons diagnosed with gastric cancers—rates of which are higher among Hispanic and non-Hispanic non-white populations[24,25]—among registries with higher proportions of non-Hispanic Asian or Pacific Islander and Hispanic patients.[26] In general, higher proportions of SEER cause-specific cause of death were missing for non-Hispanic Asian or Pacific Islander patients and Hispanic patients. One factor contributing to higher proportions of missing cause of death among Hispanic and non-Hispanic Asian or Pacific Islander patients may be related to NDI scoring (ie, that these populations have, on average, lower linkage match scores for which NDI will return vital status but not cause of death).

Missing or inaccurate Social Security number may be the underling driver of many of the cause of death missingness patterns that we see in these analyses. Among Hispanic[22] and non-Hispanic Asian or Pacific Islander patients (personal communication with Dr. Scarlett Lin Gomez, March 2022), the distribution of follow-up source central differs from that of non-Hispanic White patients, which may be due to difficulty in linking to death certificate or patient emigration. In other words, updates are made to vital status from sources that do not include cause of death, such as hospital registrars. The relatively high proportion of missing cause of death for specific primary sites (in particular, leukemia, lymphoma, and Hodgkin lymphoma) may be the result of those cancers being more frequently reported from pathology sources alone, which typically do not include Social Security number, resulting in lower matches to death data.

We also noted large differences in cause of death missingness within registry by diagnosis year. These observations speak to the dynamic nature of missingness for this field and the importance of examining missingness within registry by year and at each analysis. Inadvertently including data with a high proportion of missingness for a given year based on a low percentage missingness for all years, especially when presenting results by diagnosis year or period, could yield biased or otherwise misleading results.

One important limitation of these analyses is that by changing the method of including tumors that did not have cause of death information (excluded, dead from the cancer under study, or censored), survival estimates were not calculated using an identical tumor set, potentially biasing our comparisons. However, these calculations represent what would happen under real-life circumstances of using these data in SEER*Stat. Thus, we felt that these comparisons were appropriate. Additionally, we attempted to evaluate cause of death missingness by follow-up source central. We found that follow-up source central did not reliably capture the data source of follow-up information. For example, although linkage with the Social Security Administration – Service for Epidemiological Researchers (SSA-SER) data does not provide cause of death information, 14.7% of tumors with follow-up source central listed as SSA-SER were reported as having died of their cancer (Table 5).

Based on this evaluation of cause-specific cause of death, we recommend that the following guidelines be implemented by anyone conducting cause-specific cause of death analyses, in particular researchers and others using CiNA data:

1. Registry-specific data should be excluded from cause-specific survival calculations if >3% of cases are missing cause-specific cause of death for that registry.
2. Registry-specific data should be excluded from cause-specific survival if >10% of cases are missing cause-specific cause of death for a single year of data for that registry.
3. Because cause-specific cause of death missingness varies by primary site and race/ethnicity, researchers and others using these data should apply the above rules to the data used for their specific research questions, including analyses by the data strata of interest to their research.

Researchers may need to exclude data from additional registries depending on their study population of interest. Accordingly, researchers conducting analyses in SEER*Stat should also conduct sensitivity analyses to evaluate the impact of any missingness by registry or other subcategory on the survival estimates using these 3 options for classifying cause of death: exclusion, censoring, or dead from the cancer. Researchers conducting analyses outside of SEER*Stat should consider multiple imputation as a potential solution to missing cause-specific cause of death,

Table 4. Differences in Survival Estimates by Method and Race/Ethnicity Among Registries with < 3% Missing/Unknown Cause-Specific Cause of Death, Diagnosis Years 2011–2017

NAACCR registry number	*Absolute percentage difference in estimates*																	
	Excluded—dead from cancer						Dead from cancer—censored						Excluded—censored					
	NHW	NHB	NAIAN	NAPI	H	NU	NHW	NHB	NAIAN	NAPI	H	NU	NHW	NHB	NAIAN	NAPI	H	NU
Total	0.29	0.34	0.33	0.89	0.86	0.59	0.34	0.41	0.39	1.05	1.00	0.61	0.05	0.07	0.06	0.16	0.14	0.02
13	0.29	1.09	0.19	2.84	1.17	0.00	0.36	1.24	0.25	3.50	1.36	0.00	0.07	0.15	0.06	0.65	0.18	0.00
4	0.48	0.67	0.81	0.98	1.35	0.29	0.58	0.80	0.99	1.20	1.60	0.33	0.10	0.14	0.18	0.21	0.24	0.03
1	0.33	0.68	0.60	0.96	0.89	0.72	0.40	0.80	0.74	1.14	1.01	0.73	0.07	0.13	0.14	0.17	0.11	0.01
27	0.26	0.33	0.31	0.95	1.08	0.59	0.31	0.41	0.35	1.13	1.26	0.59	0.05	0.07	0.05	0.18	0.18	0.01
41	0.33	0.65	0.32	1.02	0.57	0.26	0.39	0.77	0.37	1.17	0.67	0.27	0.06	0.12	0.05	0.15	0.09	0.00
3	0.44	0.68	0.00	1.23	1.31	1.00	0.52	0.79	0.00	1.38	1.52	1.04	0.08	0.12	0.00	0.15	0.20	0.03
5	0.22	0.34	0.00	0.34	0.77	2.25	0.27	0.40	0.00	0.39	0.86	2.45	0.05	0.07	0.00	0.05	0.09	0.19
30	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.12	0.29	0.00	1.37	0.89	0.41	0.14	0.35	0.00	1.58	0.99	0.42	0.02	0.05	0.00	0.21	0.10	0.01
33	0.23	0.34	0.00	0.41	0.19	0.70	0.28	0.38	0.00	0.51	0.23	0.72	0.05	0.05	0.00	0.10	0.04	0.02
8	0.05	0.64	0.00	0.51	0.85	0.00	0.06	0.74	0.00	0.61	1.00	0.00	0.01	0.10	0.00	0.10	0.15	0.00
2	0.05	0.23	0.00	1.28	0.73	0.00	0.06	0.27	0.00	1.50	0.82	0.00	0.01	0.04	0.00	0.22	0.10	0.00
12	0.23	0.35	1.06	0.95	0.93	0.30	0.29	0.44	1.26	1.11	1.08	0.30	0.06	0.08	0.20	0.16	0.14	0.00
18	0.21	0.32	0.43	1.25	1.24	1.60	0.24	0.38	0.58	1.47	1.45	1.69	0.04	0.07	0.15	0.22	0.22	0.09
34	0.19	0.64	0.26	0.35	0.34	0.00	0.21	0.70	0.28	0.35	0.35	0.00	0.03	0.07	0.02	0.00	0.01	0.00
461	0.15	0.34	0.15	0.63	0.74	1.06	0.17	0.41	0.16	0.77	0.86	1.07	0.03	0.07	0.02	0.14	0.12	0.01
36	0.12	0.74	0.18	1.37	1.32	1.08	0.14	0.87	0.21	1.60	1.51	1.11	0.02	0.13	0.03	0.23	0.20	0.02
10	0.04	0.11	0.00	1.08	0.93	0.00	0.05	0.13	0.00	1.29	1.06	0.00	0.01	0.03	0.00	0.21	0.13	0.00
47	0.07	0.16	0.34	1.02	0.33	0.70	0.08	0.19	0.42	1.17	0.39	0.79	0.01	0.03	0.08	0.15	0.06	0.09
28	0.10	0.00	0.06	1.21	0.52	0.00	0.13	0.00	0.08	1.50	0.64	0.00	0.03	0.00	0.02	0.29	0.12	0.00
15	0.29	0.42	0.33	2.00	1.18	0.74	0.35	0.51	0.36	2.35	1.39	0.75	0.06	0.09	0.03	0.35	0.21	0.01
24	0.09	0.35	0.00	1.74	0.30	0.31	0.10	0.40	0.00	2.00	0.33	0.34	0.01	0.05	0.00	0.26	0.03	0.03
42	0.29	0.31	0.58	0.71	0.55	0.60	0.35	0.36	0.68	0.82	0.66	0.68	0.06	0.05	0.10	0.11	0.11	0.08
26	0.20	0.41	0.28	0.81	0.95	1.28	0.24	0.49	0.32	0.94	1.11	1.33	0.03	0.08	0.04	0.13	0.16	0.06
7	0.46	0.75	0.70	4.90	0.00	0.84	0.54	0.75	0.80	5.14	0.00	0.94	0.07	0.00	0.11	0.24	0.00	0.10
49	0.14	0.18	0.50	0.95	0.93	0.44	0.16	0.22	0.53	1.09	1.09	0.46	0.02	0.03	0.04	0.14	0.17	0.01
17	0.13	0.18	0.11	0.79	1.09	0.53	0.15	0.21	0.13	0.95	1.24	0.54	0.02	0.03	0.02	0.16	0.16	0.01
9	0.18	0.28	0.22	0.65	0.70	0.35	0.21	0.33	0.25	0.75	0.80	0.36	0.03	0.05	0.04	0.10	0.09	0.01
45	0.03	0.03	0.00	0.14	0.32	0.09	0.04	0.04	0.00	0.15	0.34	0.09	0.01	0.01	0.00	0.01	0.02	0.00
16	0.03	0.00	0.04	0.00	0.73	0.00	0.04	0.00	0.05	0.00	0.83	0.00	0.01	0.00	0.01	0.00	0.10	0.00
29	0.11	0.19	0.99	0.63	1.07	0.69	0.13	0.22	1.12	0.75	1.20	0.76	0.02	0.03	0.14	0.12	0.13	0.07
21	0.23	0.37	0.49	1.17	0.83	1.98	0.27	0.43	0.57	1.34	0.96	2.03	0.04	0.06	0.08	0.17	0.13	0.05
25	0.08	0.59	0.00	0.74	0.91	0.00	0.09	0.66	0.00	0.87	1.05	0.00	0.01	0.07	0.00	0.13	0.15	0.00
31	0.09	0.08	0.11	0.13	0.22	0.37	0.10	0.08	0.11	0.13	0.23	0.37	0.01	0.00	0.00	0.01	0.02	0.00
32	0.04	0.09	0.00	0.90	0.31	0.35	0.05	0.10	0.00	1.08	0.34	0.38	0.01	0.02	0.00	0.18	0.04	0.03
20	0.29	0.19	0.45	0.33	0.31	1.02	0.32	0.22	0.49	0.37	0.34	1.10	0.03	0.03	0.04	0.04	0.03	0.09
Median diff	0.16	0.33	0.19	0.95	0.80	0.43	0.19	0.39	0.23	1.10	0.91	0.44	0.03	0.05	0.02	0.15	0.12	0.01

H, Hispanic (all races); diff, difference; NAIAN, non-Hispanic American Indian/Alaska Native; NAPI, non-Hispanic Asian or Pacific Islander; NHB, non-Hispanic Black; NHW, non-Hispanic white; NU, non-Hispanic unknown race.

Table 5. Cause of Death Missingness by Follow-up Source Central for Deceased Patients Diagnosed While Residents of Registries with <3% Missing Cause of Death, 2011–2017 Diagnosis Years, All Sites

	Alive or dead of other cause		*Dead (attributable to this cancer)*		*Dead (missing/ unknown COD)*		*Total*
Follow-up source central	**Count**	**Row %**	**Count**	**Row %**	**Count**	**Row %**	**Count**
Follow-up not performed for this patient	905	8.70	6,989	67.20	2,506	24.10	10,400
Medicare/Medicaid File	0	0.00	3	5.60	51	94.40	54
Center for Medicare and Medicaid Services (formerly Health Care Finance Administration [HCFA])	4	0.90	17	4.00	405	95.10	426
Department of Motor Vehicle Registration	3	16.70	6	33.30	9	50.00	18
National Death Index (NDI)	23,197	26.30	64,683	73.40	205	0.20	88,088
State death tape/death certificate file	525,157	23.30	1,730,082	76.70	1,627	0.10	2,256,873
County/municipality death tape/ death certificate file	9	1.70	508	97.90	2	0.40	519
Social Security Administration Death Master File	34,071	20.10	130,498	76.80	5,321	3.10	169,890
Hospital discharge data	135	8.70	1,302	84.40	106	6.90	1,543
Health maintenance organization (HMO) file	5	3.90	9	7.00	114	89.10	128
Social Security epidemiological vital status data	163	5.90	406	14.70	2,197	79.40	2,766
Voter registration file	6	25.00	8	33.30	10	41.70	24
Linkages, NOS	10,894	22.20	37,520	76.30	746	1.50	49,160
Hospitals and treatment facilities	7,179	18.60	26,464	68.50	4,998	12.90	38,642
Physicians	388	7.10	4,690	85.30	423	7.70	5,501
Patient	1	0.70	4	2.70	141	96.60	146
Central or regional cancer registry	288	18.20	1,129	71.40	165	10.40	1,582
Other	161	5.30	1,559	51.20	1,326	43.50	3,046
Blank(s)	26	11.90	184	84.00	9	4.10	219
Total	602,592	22.90	2,006,061	76.30	20,361	0.80	2,629,025

COD, cause of death; NOS, not otherwise specified.

with special consideration for analyses involving registries impacted by legislation- or registry operations–related reason for high missingness.

These analyses underscore that registries differ in their interpretation or have different policies governing cause of death release. Cause of death ascertained from state death records may not be releasable by central cancer registries per agreements with state vital statistics departments. Cause of death information is, however, used to calculate SEER cause-specific cause of death, which may be able to be released by central cancer registries in the absence of specific cause of death information. Additionally, as described in the data release guidelines published by NAACCR,[27] data on fact, date, and cause of death identified through NDI linkages may be released to approved researchers after review and approval by the cancer registry provided that the registry annually provides NDI with information describing the release of these data (ie, researcher name, organization, study title, date). Release of NDI fact, date, and cause of death may be included in annual data submissions to NPCR, the SEER Program, and NAACCR.

The results of these analyses also underscore the importance of data processing sequence in the annual data submissions. Specifically, if death linkage is conducted prior to a particular case being reported to the registry (eg, late reporting of interstate data or delayed reporting by hospitals), vital status may not be appropriately recorded for that patient. Central registries should be particularly cognizant of the potential need to reconduct death linkages for these new cases.

Conclusion

This paper aims to establish a standard for when registry data is or is not fit for use for cause-specific survival with the ultimate goal of encouraging registries to improve their data quality. To this end, we have established

a recommended cut point of <3% missing/unknown SEER cause-specific cause of death by registry and/or any strata for which cause-specific survival is reported. We have also established that any registry with <3 % overall missing/unknown SEER cause-specific cause of death but >10% missing/unknown SEER cause-specific cause of death for 1 or more individual diagnosis years should also be excluded from cause-specific survival analyses in SEER*Stat. This 3% standard is a reasonable request of most registries, as most US registries already met the standard prior to its quantification. This cause-specific cause of death fit-for-use criterion is a direct analogue to how other fit-for-use metrics are applied to CiNA data products, and maintaining a similar approach to cause-specific survival is sensible for CiNA data products and surveillance publications.

This paper also serves to call researchers' attention to how missingness in cancer surveillance data is differential and likely to impact survival estimates. In the absence of multiple imputation or other more advanced statistical techniques, we recommend that researchers working with any subnational data set or who intend to present survival estimates by subpopulations exclude data per the above criteria. At minimum, we recommend conducting sensitivity analyses as described.

References

1. Sherman R, Firth R, Kahl A, et al, eds. *Cancer in North America: 2015–2019: Volume One: Combined Cancer Incidence for the United States, Canada and North America*. North American Association of Central Cancer Registries, Inc; 2022.
2. Sherman R, Firth R, Kahl A, et al, eds. *Cancer in North America: 2015–2019: Volume Two: Registry-Specific Cancer Incidence in the United States and Canada*. North American Association of Central Cancer Registries, Inc; 2022.
3. Sherman R, Firth R, Kahl A, et al, eds. *Cancer In North America, 2015–2019: Volume Three: Registry-Specific Cancer Mortality in the United States and Canada*. North American Association of Central Cancer Registries, Inc; 2022.
4. Johnson CJ, Wilson R, Mariotto A, et al, eds. *Cancer in North America: 2015–2019: Volume Four: Cancer Survival in the United States and Canada 2012–2018*. North American Association of Central Cancer Registries, Inc; 2022.
5. Johnson CJ, Wilson R, Mariotto A, et al, eds. *Cancer in North America: 2015–2019: Volume Five: Cancer Prevalence in the United States and Canada 2009–2018*. North American Association of Central Cancer Registries, Inc; 2022.
6. Weir HK, Tucker TC. NAACCR registry certification. In: Menck HR, Deapen D, Phillips JL, Tucker TC, eds. *Central Cancer Registries: Design, Management and Use*. 2nd ed. Kendall/Hunt Publishing Company; 2007:223-236.
7. Perme MP, Stare J, Esteve J. On estimation in relative survival. *Biometrics*. 2012;68(1):113-120.
8. Forjaz de Lacerda G, Howlader N, Mariotto AB. Differences in cancer survival with relative versus cause-specific approaches: an update using more accurate life tables. *Cancer Epidemiol Biomarkers Prev*. 2019;28(9):1544-1551.
9. Wissing MD, Greenwald ZR, Franco EL. Improving the reporting of cancer-specific mortality and survival in research using cancer registry data. *Cancer Epidemiol*. 2019;59:232-235.
10. Mariotto AB, Zou Z, Johnson CJ, Scoppa S, Weir HK, Huang B. Geographical, racial and socio-economic variation in life expectancy in the US and their impact on cancer relative survival. *PLoS One*. 2018;13(7):e0201034.
11. Rachet B, Maringe C, Woods LM, Ellis L, Spika D, Allemani C. Multivariable flexible modelling for estimating complete, smoothed life tables for sub-national populations. *BMC Public Health*. 2015;15:1240.
12. Spika D, Bannon F, Bonaventure A, et al. Life tables for global surveillance of cancer survival (the CONCORD programme): data sources and methods. *BMC Cancer*. 2017;17(1):159.
13. Howlader N, Ries LA, Mariotto AB, Reichman ME, Ruhl J, Cronin KA. Improved estimates of cancer-specific survival rates from population-based data. *J Natl Cancer Inst*. 2010;102(20):1584-1598.
14. Lauseker M, Zu Eulenburg C. Analysis of cause of death: competing risks or progressive illness-death model? *Biom J*. 2019;61(2):264-274.
15. Forjaz G, Howlader N, Scoppa S, Johnson CJ, Mariotto AB. Impact of including second and later cancers in cause-specific survival estimates using population-based registry data. *Cancer*. 2022;128(3):547-557.
16. Nur U, Shack LG, Rachet B, Carpenter JR, Coleman MP. Modelling relative survival in the presence of incomplete data: a tutorial. *Int J Epidemiol*. 2010;39(1):118-128.
17. Carpenter JR, Kenward MG. *Multiple Imputation and Its Application*. John Wiley & Sons; 2013.
18. Binder N, Schumacher M. Missing information caused by death leads to bias in relative risk estimates. *J Clin Epidemiol*. 2014;67(10):1111-1120.
19. SEER*Stat Database: NAACCR Incidence Data - CiNA Production File, 1995-2018, for U.S. and CDN - Survival & Prevalence (which includes data from CDC's National Program of Cancer Registries (NPCR), CCCR's Provincial and Territorial Registries, and the NCI's Surveillance, Epidemiology and End Results (SEER) Registries), certified by the North American Association of Central Cancer Registries (NAACCR) as meeting high-quality incidence data standards for the specified time periods, submitted December 2020. In: 2021.
20. Thornton ML, ed. *Standards for Cancer Registries Volume II: Data Standards and Data Dictionary*. 24th ed. North American Association of Central Cancer Registries; 2022.
21. SEER*Stat software [computer program]. Version 8.4.0.1. www.seer.cancer.gov/seerstat
22. Pinheiro PS, Morris CR, Liu L, Bungum TJ, Altekruse SF. The impact of follow-up type and missed deaths on population-based cancer survival studies for Hispanics and Asians. *J Natl Cancer Inst Monogr*. 2014;2014(49):210-217.
23. Morawski BM, Qiao B, Coyle L, Rycroft RK, Schymura MJ, Johnson CJ. Impact of linkage to the Social Security Administration on follow-up completeness and cancer relative survival estimates in 2 new SEER registries: 2000–2016 diagnosis years. *J Registry Manag*. 2020;47(2):37-47.
24. Haile RW, John EM, Levine AJ, et al. A review of cancer in U.S. Hispanic populations. *Cancer Prev Res (Phila)*. 2012;5(2):150-163.
25. Everhart JE, Kruszon-Moran D, Perez-Perez GI, Tralka TS, McQuillan G. Seroprevalence and ethnic differences inHelicobacter pyloriInfection among adults in the United States. *J Infect Dis*. 2000;181(4):1359-1363.
26. Zavala VA, Bracci PM, Carethers JM, et al. Cancer health disparities in racial/ethnic minorities in the United States. *Br J Cancer*. 2021;124(2):315-332.
27. *Linking Central Cancer Registry Data with National Death Index (NDI) Data*. NAACCR; 2021. **https://www.naaccr.org/wp-content/uploads/2021/04/NDI-Factsheet-for-NAACCR_Final_Updated-4.29.21.pdf**

Short Report

Prostate Cancer Trends in Montana

Heather J. N. Zimmerman, MPH [a]

Abstract: Prostate cancer is the most common cancer among men in Montana and the second-most-common cause of cancer deaths. In 2014, prostate cancer incidence in Montana started increasing significantly, even as incidence in the United States overall stayed about the same. The increased incidence was not accompanied by an increase in prostate cancer mortality. Trends in local stage incidence and incidence among men aged 65 to 79 years mirrored the trends in overall prostate cancer incidence and suggest that changes are due to screening behavior. However, it is difficult to determine what may have caused increased screening among Montana men since 2014. Monitoring prostate cancer incidence and mortality is an important tool in determining if there is a change in prostate cancer disease burden or in overdiagnosis, and informs planning for possible public health intervention.

Key words: *prostate cancer, Montana*

Background

Prostate cancer is the most common type of cancer among men, with about 990 new cases in Montana each year.[1] Survival is very high among prostate cancer patients; 96% of patients survive for at least 5 years after diagnosis.[2] But prostate cancer is still the second-most-common cause of cancer-related deaths among men after lung cancer. About 140 Montana men die from prostate cancer each year.[3]

Prostate cancer can be detected through screening. The prostate-specific antigen (PSA) test, a blood test that measures levels of proteins produced by prostate cells, along with a digital rectal examination, can provide an early signal of prostate cancer. Screening behavior has a large impact on incidence, since many prostate cancers never cause symptoms and, without screening, would never be diagnosed. In autopsy studies of US White or European men who died of other causes, prostate cancer was found in 22% of men aged 50 to 59 years, 29% of men aged 60 to 69 years, 36% of men aged 70 to 79 years, and 47% of men aged 80 years or older.[4] Among US Black men, prostate cancer was even more common, found in 46% of men aged 50 to 59 years, 47% of men aged 60 to 69 years, and 50% of men aged 70 to 79 years.[4]

PSA testing was approved by the US Food and Drug Administration to screen for prostate cancer in 1994 and many organizations recommended annual screening in men aged 50 years and older until about 2008.[5] However, screening recommendations have varied widely between organizations and over time. PSA testing may be beneficial to some men, especially if they are at high risk for prostate cancer, but there are also some potential problems with routine PSA testing. PSA levels may be high because of other conditions unrelated to cancer.[5] These false positive results may lead to unnecessary biopsies and potential adverse effects from the biopsy. Additionally, many prostate cancers will never become symptomatic and do not need to be treated. When these cases are detected and treated, it is called *overdiagnosis*. Follow-up of large randomized trials suggests that 20% to 50% of men diagnosed with prostate cancer through screening may be overdiagnosed.[6] Because of the potential harms from false positives and overdiagnosis, in 2012, the US Preventive Services Task Force (USPSTF) concluded that the benefits of screening did not outweigh the harms and recommended against routine screening for average-risk men.[7] In 2018, USPSTF updated their recommendation to state that men aged 55 to 69 years should discuss the potential risks and benefits of screening with their health care provider before undergoing screening.[7] USPSTF still recommends against prostate cancer screening for men aged 70 years or older. Many other organizations have similar recommendations for prostate cancer. Like USPSTF, the American Cancer Society recommends shared decision-making, but recommends starting that discussion at age 50 years for average-risk men; at age 45 years for African American men and men with a first-degree relative diagnosed with prostate cancer before age 65 years; and at age 40 years for men with more than 1 first-degree relative diagnosed with prostate cancer before the age of 65 years.[8]

This report explores the trends in prostate cancer incidence and mortality in Montana and the extent to which observed changes are due to changes in screening behavior.

Methods

Data from the Montana Central Tumor Registry and Montana death certificates were used to calculate age-adjusted prostate cancer incidence and mortality rates among Montana men from 1995 to 2019. Montana rates were compared to rates in the United States overall using the US Cancer Statistics public use data set for all years available at the time of analysis (1999–2018).[9] Trends in incidence and mortality rates in Montana and the United States overall were examined using Joinpoint statistical software version

[a] Montana Department of Public Health and Human Services.

Address correspondence to Heather J. N. Zimmerman, MPH. Email: hzimmerman@mt.gov.

This publication was supported by the cooperative agreement number DP17-1701 from the Centers for Disease Control and Prevention (CDC). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the CDC.

4.9.0.0 developed by IMS, Inc under contract to the National Cancer Institute. Incidence rates were also calculated by age at diagnosis and by stage at diagnosis among Montana men.

Results

Prostate cancer incidence in Montana remained about the same (around 190 new cases per 100,000 men) from 1999 to 2004 (Figure 1). Incidence then started decreasing in Montana, with an annual percent change (APC) of –5.5% from 2004 to 2014. Prostate cancer trends in the United States overall were very similar as to Montana, with rates remaining around 170 per 100,000 from 1999 to 2007, followed by a significant decrease from 2007 to 2014 (APC, –6.6%). After 2014, the trends in Montana and the United States diverged, with Montana rates increasing significantly (APC, 5.2%) while rates in the United States stayed around 105 per 100,000 from 2014 to 2019. Prostate cancer mortality rates were about the same in Montana and the United States overall and have been steadily declining since 1999.

Figure 1. Prostate Cancer Incidence (New Cases) and Mortality (Deaths) in Montana and the United States, 1999–2019

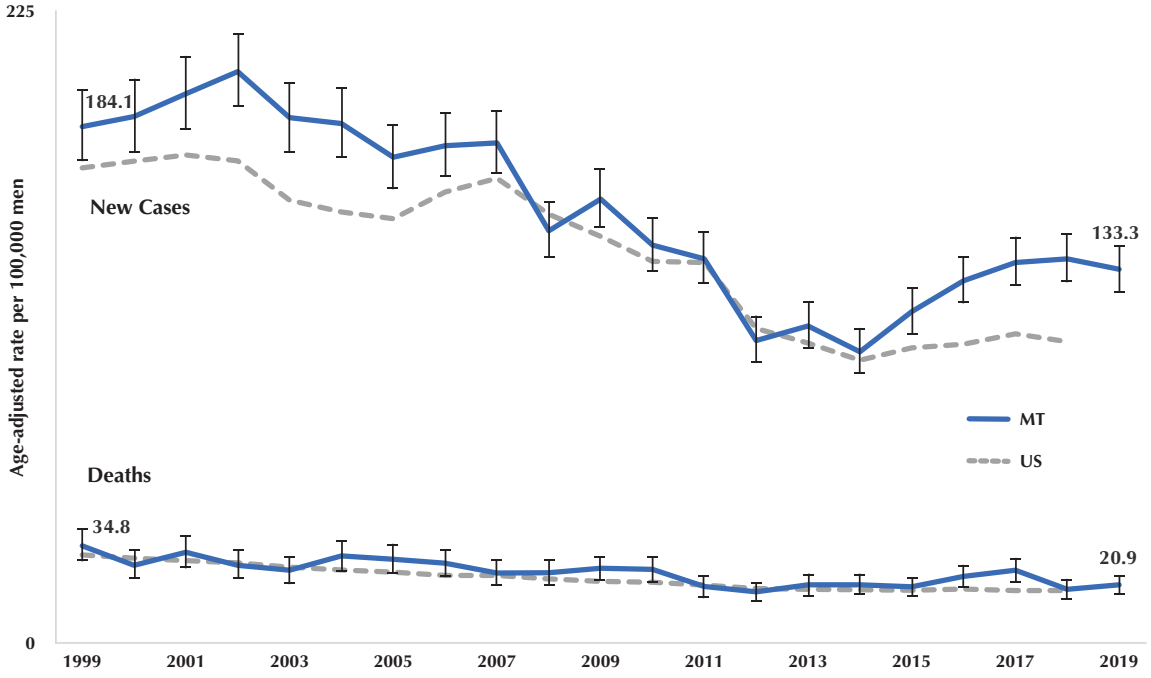


Figure 2. Age-Adjusted Prostate Cancer Incidence Trends by Stage at Diagnosis, Montana, 1995–2019

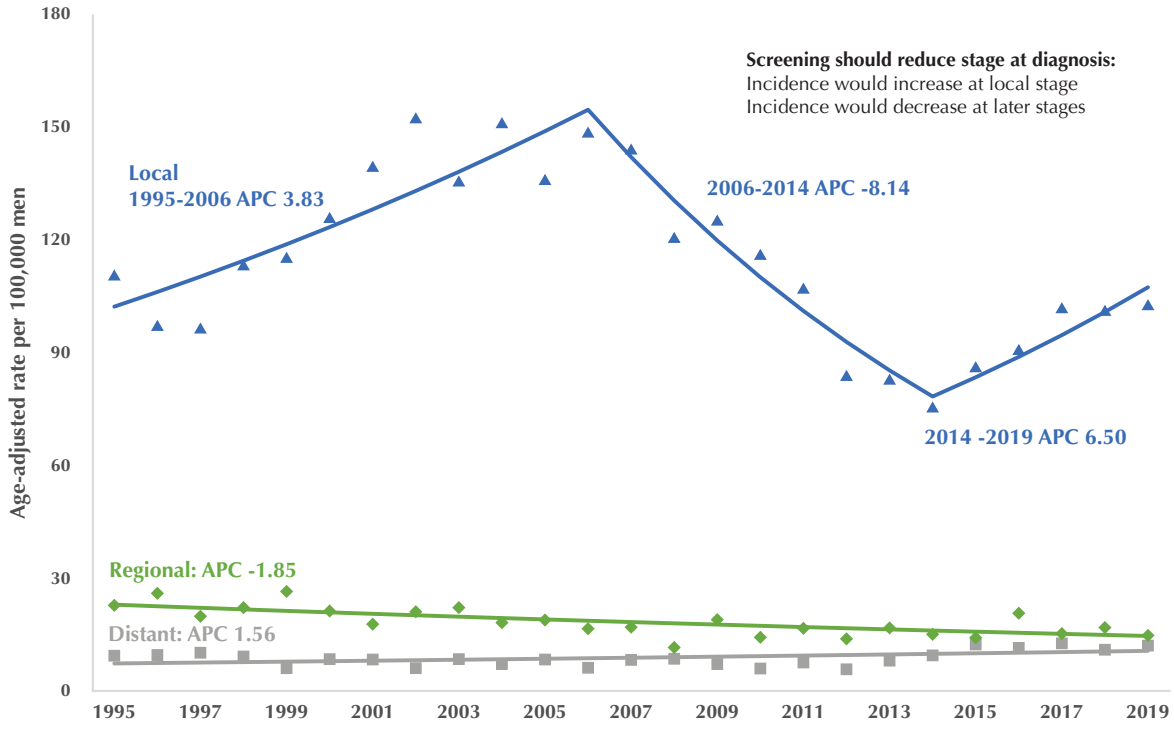
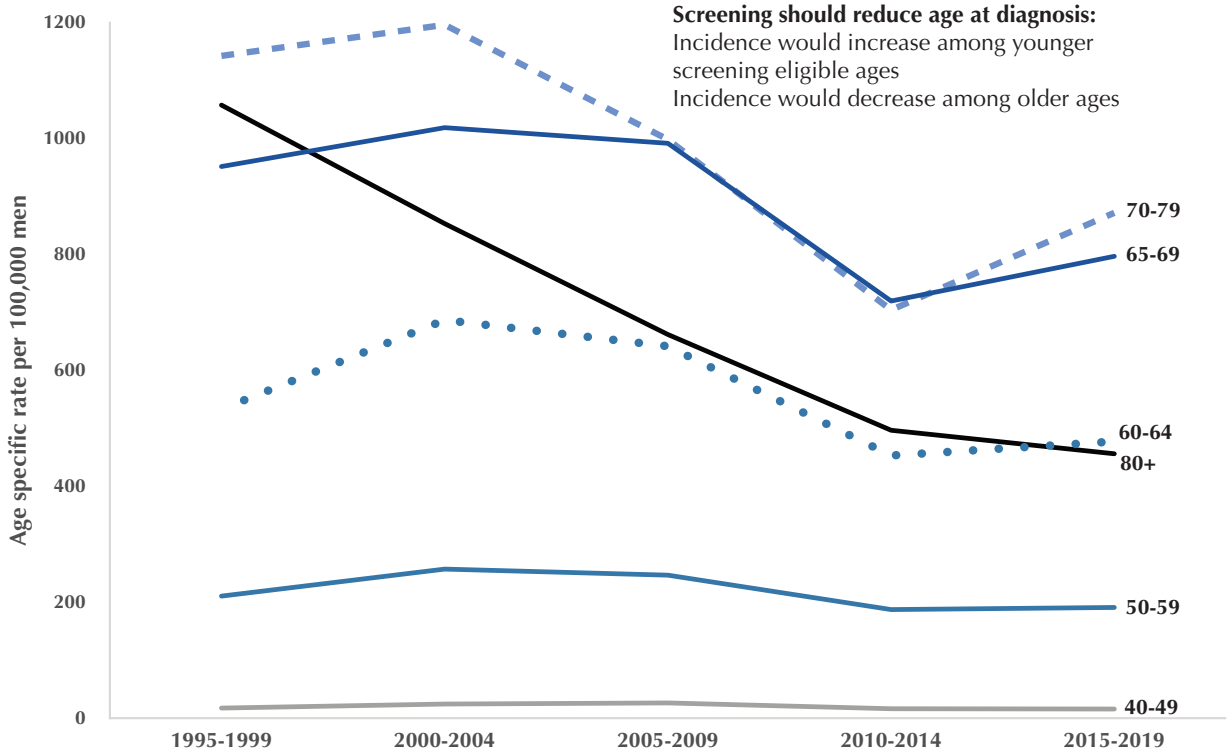


Figure 3. Age-Specific Prostate Cancer Incidence Trends, Montana, 1995–2019



Prostate cancer incidence trends in Montana vary by stage at diagnosis. Local stage incidence increased from 1995 to 2006, with an APC of almost 4%, and then decreased from 2006 to 2014 (APC, –8.1%) and rose again from 2014 to 2019 (APC, 6.5%) (Figure 2). From 1995 to 2019, regional stage incidence decreased (APC, –1.9%) and distant stage incidence increased (APC, 1.6%). Prostate cancer incidence trends were also different based on age at diagnosis. Incidence among younger men (aged 40 to 59 years) was low and stayed about the same from 1995 to 2019 (Figure 3). Incidence among men aged 60 to 79 years increased slightly from 1995 to 2000, decreased from 2000 to 2014, and then increased again from 2014 to 2019, similar to the overall statewide rate trends. Among men aged 80 years and older, incidence decreased dramatically from 1995 to 2019. Men aged 65 to 79 years had the highest rates from 2000 forward.

Discussion

Effective cancer screening programs should increase the diagnosis of local stage disease and decrease the diagnosis of late-stage disease. Screening should also reduce age at diagnosis so that incidence would increase among younger age groups who are eligible for screening and decrease among older age groups. The prostate cancer incidence trends seen in Montana since 1995 are consistent with changes caused by screening behavior, as most of the change was in local stage disease among men aged 65 to 79 years. They are consistent with national trends and can be attributed to changes in screening recommendations until 2014. Even after 2014, prostate cancer incidence trends may be due to increased screening among Montana men despite screening recommendations staying the same. More men may be getting screened in Montana because of increased health insurance coverage starting in 2016. Montana Medicaid expanded eligibility as part of the Affordable Care Act on January 1, 2016, and almost 95,000 people gained coverage through the program as of 2019.[10] This increased insurance coverage did lead to a measurable rise in the utilization of other preventive health services.[10] However, there may be other factors contributing to the increased incidence.

Higher risk of prostate cancer is associated with older age, African American race, and having a father or brother with prostate cancer.[8] Some behavioral risk factors have been suggested but there is not clear evidence to show they increase prostate cancer risk. As Montana has not had a significant demographic shift since 2014, it is unlikely that the increased incidence is related to a change in prostate cancer risk among Montana men and suggests that there is not a need for any public health interventions to prevent prostate cancer or increase screening.

Prostate cancer mortality in Montana has not increased since 2014 and remains about the same as in the United States overall. In fact, prostate cancer mortality continues to decrease both nationally and in Montana. This is promising and suggests that the increased incidence is not necessarily problematic. However, it may be an indication of overdiagnosis. The Montana Central Tumor Registry will continue to monitor prostate cancer incidence and mortality trends to determine if intervention is needed to mitigate the risks of overdiagnosis.

References

1. Montana Department of Public Health and Human Services. Montana Central Tumor Registry, 2015–2019.
2. Centers for Disease Control and Prevention. *U.S. Cancer Statistics Prostate Cancer Stat Bite*. US Department of Health and Human Services; 2022.
3. Montana Department of Public Health and Human Services. Montana Vital Statistics Office, death certificate data, 2015–2019.
4. Jahn JL, Giovannucci EL, Stampfer MJ. The high prevalence of undiagnosed prostate cancer at autopsy: implications for epidemiology and treatment of prostate cancer in the prostate-specific antigen-era. *Int J Cancer*. 2015;137(12):2795-2802.
5. Prostate-specific antigen (PSA) test. National Cancer Institute website. Reviewed March 11, 2022. Accessed June 2022. **https://www.cancer.gov/types/prostate/psa-fact-sheet**
6. Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow J. *Prostate-Specific Antigen–Based Screening for Prostate Cancer: A Systematic Evidence Review for the US Preventive Services Task Force*. Evidence Synthesis No. 154. AHRQ Publication No. 17-05229-EF-1. Agency for Healthcare Research and Quality; 2018.
7. Final recommendation statement: prostate cancer: screening. US Preventive Services Task Force website. Published May 8, 2018. Accessed June 2022. **https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/prostate-cancer-screening**
8. American Cancer Society recommendations for prostate cancer early detection. American Cancer Society website. Updated April 23, 2021. Accessed June 2022. **https://www.cancer.org/cancer/prostate-cancer/early-detection/acs-recommendations.html**
9. National Program of Cancer Registries and Surveillance, Epidemiology and End Results Program SEER*Stat Database: NPCR and SEER Incidence —U.S. Cancer Statistics Public Use Research Database, 2020 Submission (2001–2018). United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released June 2021. **www.cdc.gov/cancer/uscs/public-use**
10. Rasmussen R. *Analysis of the Impact of Medicaid Expansion on Montana*. Navigant; 2019. **https://mtha.org/wp-content/uploads/2021/03/2019_Navigant_Analysis-of-Impact-of-Medicaid-Expansion-on-Montana.pdf**
11. Prostate cancer risk factors. American Cancer Society website. Published June 9, 2020. Accessed June 2022. **https://www.cancer.org/cancer/prostate-cancer/causes-risks-prevention/risk-factors.html**

NAACCR First-Place Poster

Participation Rates and Characteristics of Participants and Nonparticipants in Cancer Patient Contact Studies in New York State

Margaret Gates Kuliszewski[a,b]; Heather Sickler[a]; Marilyn Sango-Jordan[a]; Stephanie Saint Germain[a]; Sarah Moncrief Feliciano[a]; Nicole Vyce[a]; Maria J. Schymura[a,b]

Background

- Cancer research studies involving patient contact often use cancer registries to identify eligible patients, with a goal of recruiting a representative sample
- Low rates of participation may decrease the representativeness of the sample, introduce selection bias, and limit generalizability of study results
- We examined participation rates and characteristics of participants and nonparticipants for 6 recent patient contact studies conducted in New York State (NYS)

Methods

- Analyses included patients who were sent an initial study mailing for 1 of 6 patient contact studies before November 1, 2021 (Table 1)
- We conducted descriptive analyses to examine the percent of patients who consented to further contact or participated in each study
- We used χ^2 and Fisher's exact tests to compare demographic characteristics of participants and nonparticipants in each study
- All analyses were conducted using SAS 9.4

Results

- For studies requiring active patient consent, 25.3% of patients agreed to further contact for an intervention study vs 51.3% for a study involving a telephone interview only (Figure 1)
- For studies requiring passive patient consent, 9% to 10% of patients assented to further contact and 75% to 79% passively consented after 4 weeks (Figure 1)
- For 2 prostate cancer studies involving direct enrollment by survey completion, 42.9% and 16.5% participated in each study as of February 2022 (Figure 2)
- There were statistically significant differences between participants and nonparticipants in age at diagnosis, race, and region/county at diagnosis for 1 or more patient contact studies:
  - Thyroid cancer patients who consented to further contact were more likely to be White and less likely to be Asian, Black, or other/unknown race ($P < .001$); no differences were observed for diagnosis year, sex, or age category

Table 1. Patient Contact Studies Included in Analyses

Cancer type	*Diagnosis years*	*Physician consent*	*Patient Consent*	*Study Procedures*
Thyroid	2011–2021	Passive	Active	Telephone interview
Bladder	2018–2019	Passive	Active	Intervention, biospecimen
Breast (female)	2019–2020	Passive	Passive	Surveys, biospecimen
Ovarian	2019–2020	Passive	Passive	Survey, optional biospecimen
Prostate	2020–2021	None	Survey Completion	Survey, optional HIPAA form
Prostate	2016	None	Survey Completion	Survey, optional biospecimen

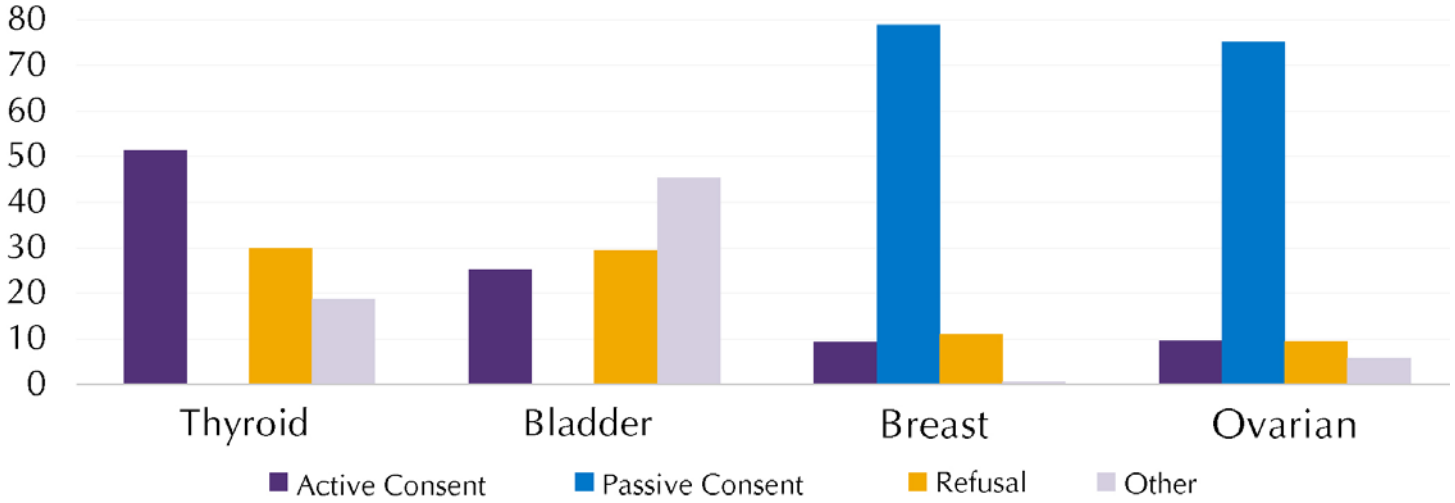
Note: Passive consent provides the opportunity to decline further contact about a study; active consent requires assent to further contact about a study.

[a]New York State Cancer Registry, New York State Department of Health, Albany, New York. [b]Department of Epidemiology and Biostatistics, University at Albany School of Public Health, Rensselaer, New York.

This content was originally presented as a poster at the NAACCR 2022 Summer Forum, June 14–16, 2022.

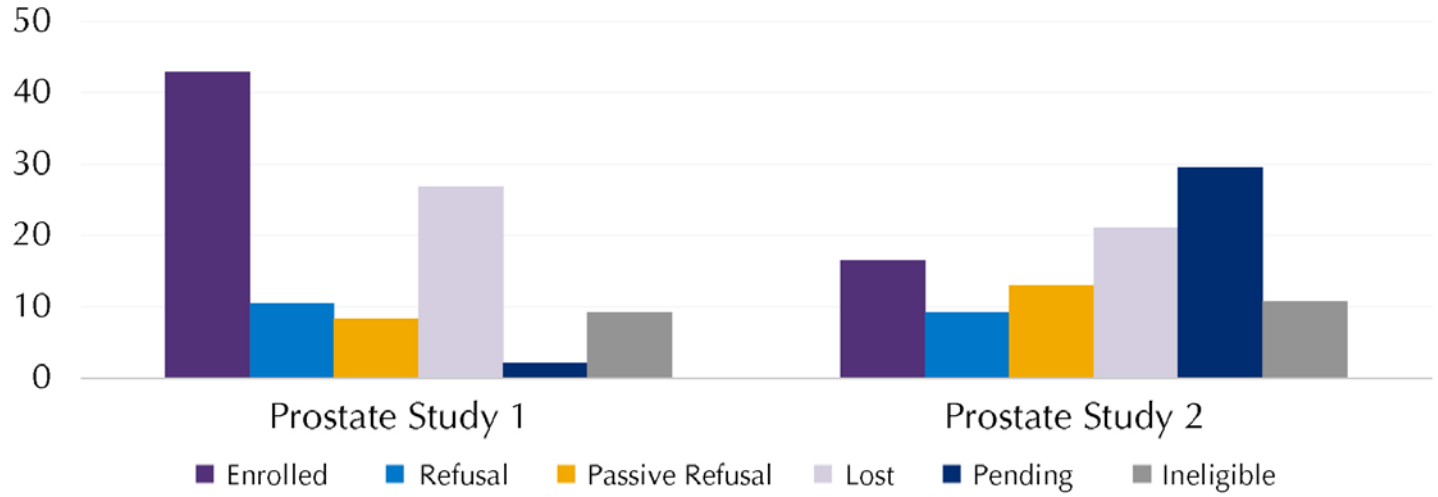
This work was supported in part by cooperative agreement 6NU58DP006309 awarded to the New York State Department of Health by the Centers for Disease Control and Prevention and by Contract 75N91018D00005 (Task Order 75N91018F00001) from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services. The studies referenced were funded by subcontracts via awards from the National Cancer Institute, the Breast Cancer Research Foundation, the Patient-Centered Outcomes Research Institute, and the Prostate Cancer Foundation, as well as funding from United BioSource, LLC.

Figure 1. Participation Rates for Studies Requesting Patient Consent for Further Contact



Note: "Other" category includes patients who are ineligible, unable to be located, or pending.

Figure 2. Participation Rates for Studies Involving Direct Enrollment of Patients



- Bladder cancer patients who consented to further contact were more likely to be ages 60–69 or 70-79 years and less likely to be <60 or ≥80 years of age (P = .02); no differences were observed for race/ethnicity, diagnosis year, or sex
- Breast cancer patients who actively consented were more likely to live outside of New York City (NYC) (P < .001), while patients who refused further contact were more likely to be 60–69 or ≥70 years of age (P = .002); no differences were observed for race or diagnosis year
- Ovarian cancer patients who actively consented were more likely to be White (P = .009) and live outside of NYC (P = .002); no differences were observed for diagnosis year or age category
- In 1 prostate cancer study, participants were more likely to be White, and in both prostate cancer studies, participants were more likely to live outside of NYC (all P < .001); no differences were observed for ethnicity or age category

Conclusion

- In recent patient contact studies conducted by the NYS Cancer Registry, participation rates varied by study design and contact procedures
  - The overall study response rate is unknown for the nonenrollment studies
- Patient self-selection may lead to a study sample that differs from the underlying population, including with respect to age, race, and region of NYS

NAACCR Second-Place Poster

Racial/Ethnic Disparities in COVID-19 Infection Among Working-Age Women with Precancerous Cervical Lesion

Mei-Chin Hsieh, PhD, CTR[a]; Tina Lefante, MPH, CTR[a]; Yong Yi, PhD[a]; Xiao-Cheng Wu, MD, MPH, CTR[a]

Key Messages

- The racial/ethnic disparities in COVID-19 infection among women with precancerous cervical lesion (PCL) was observed.
- Non-Hispanic Black (NHB) and Hispanic PCL women had a higher risk of COVID-19 infection than Non-Hispanic White (NHW) women.
- Younger working-age women had a higher risk of COVID-19 infection.
- Women residing outside of the greater New Orleans area had an increased risk of COVID-19 infection.

Background

PCL is most likely diagnosed in working-age women. In Louisiana, over 98% of PCL cases were diagnosed at age 18–65 years, and women aged 20–34 years had the highest incidence rate. Before the Omicron variant spread, COVID-19 prevalence was higher in young and middle-aged adults and minorities in the United States. Because most PCL cases occurred in the similar age group as COVID-19, this study aimed to assess the racial/ethnic disparities in COVID-19 infection on this specific population.

Objective

To assess the racial/ethnic disparities in COVID-19 infection among PCL women aged 18–65 years.

Methods

Study Cohort

Women with precancerous cervical lesion (PCL) diagnosed in 2009–2021, including CIN3, carcinoma in situ (CIS), severe dysplasia, adenocarcinoma in situ (AIS), and high-grade dysplasia for year ≥2019, were obtained from the Louisiana Tumor Registry. We included PCL women aged 18–65 years either at time of a COVID-19 diagnosis or year 2021 for COVID-free patients. PCL women who died before 2020 were excluded.

COVID-19 Data

We linked eligible PCL women with the Louisiana statewide COVID-19 2020–2021 data to identify patients with COVID-19 infection.

Variables

- Outcome variable: COVID-19 diagnosis (yes vs no).
- Race/ethnicity was categorized as NHW, NHB, Hispanic, and others.
- Other covariates included age, marital status, type of insurance, census-track level socioeconomic status (SES) quintiles (higher group indicates the higher SES), and Louisiana metropolitan status (nonmetro, greater New Orleans, and other metro).

Statistical Analysis

Descriptive statistics on covariates by COVID-19 infection status were presented and χ^1 test was used to assess the unadjusted association. Logistic regression was employed to assess the racial/ethnic differences in COVID-19 infection among working-age women with PCL disease.

Results

- Of 14,589 eligible PCL women, 11.3% were diagnosed with COVID-19 and 83% were confirmed with a positive polymerase chain reaction (PCR) test.
- NHB women had the highest percentage of COVID-19 infection (14.1%), followed by Hispanic women (11.8%) (Table 1). NHW women and women with other race/ethnicity had a similar infection rate (9.8% and 9.3%, respectively) (Table 1).
- The COVID-19 infection rate decreased as age increased, with 13.2% for PCL women aged 18–29 years and 9.6% for those aged 50–65 years.
- COVID-19 infection was higher in 2020 than in 2021 across all racial/ethnic groups (Figure 1).
- In the crude model, only NHB women showed a higher risk of COVID-19 infection than NHW women. However, after adjustment, the odds of infection for NHB women were 55% higher (95% CI, 1.38–1.75) and for Hispanic women was 30% higher (95% CI, 1.03–1.63) than NHW women (Table 2).
- Compared to those aged 50–65 years, younger age groups were more likely to have COVID-19 with an adjusted odds ratio (aOR) of 1.43 (95% CI, 1.16–1.76) for ages 18–29 years, 1.22 (95% CI, 1.03–1.45) for ages 30–39 years (Table 2).
- Additionally, compared to greater New Orleans, PCL women residing in nonmetro areas or other metro counties had a higher risk of COVID-19 infection with an aOR of 1.24 (95% CI, 1.02–1.51) and 1.17 (95% CI, 1.01–1.36), respectively (Table 2).

[a]Louisiana Tumor Registry/Epidemiology Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, Louisiana.

This content was originally presented as a poster at the NAACCR 2022 Summer Forum, June 14–16, 2022.

This work was supported in part by Louisiana State University Health Sciences Center, the Centers for Disease Control and Prevention under cooperative agreement of the National Program of Cancer Registries grant number NU58DP006332, and the National Cancer Institute's contract number HHSN2612018000071.

Conclusions

After adjusting for age and socioeconomic covariates, we found substantial variation in racial/ethnic disparities in COVID-19 infection among working-age PCL women. Other risk factors that could cause these disparities, such as comorbidities and individual behavior, need further investigation.

Table 1. Patient Characteristics Among Working-Age Women with Precancerous Cervical Lesion Diagnosed in 2009–2021 by COVID-19 Infection Status

Variables	COVID-19 Diagnosis No (N=12,945) Count (%)	Yes (N=1,644) Count (%)	p-value
Race/Ethnicity			<.0001
NHW	7952 (90.2)	860 (9.8)	
NHB	4099 (85.9)	671 (14.1)	
Hispanic	698 (88.2)	93 (11.8)	
Other	196 (90.7)	20 (9.3)	
Age[1]			0.0058
18-29	1561 (86.8)	237 (13.2)	
30-39	6388 (88.6)	821 (11.4)	
40-49	3310 (89.0)	408 (11.0)	
50-65	1686 (90.5)	178 (9.6)	
Marital status			0.1002
Married	1031 (88.0)	140 (12.0)	
Single	2178 (87.7)	307 (12.4)	
Unknown	9736 (89.1)	1197 (11.0)	
Health insurance			0.0827
P/M/O	979 (86.7)	150 (13.3)	
Medicaid	638 (89.0)	79 (11.0)	
No/unknown	11328 (88.9)	1415 (11.1)	
SES			0.0197
Group 1	2612 (87.3)	379 (12.7)	
Group 2	2576 (88.7)	328 (11.3)	
Group 3	2546 (89.7)	292 (10.3)	
Group 4	2679 (88.3)	354 (11.7)	
Group 5	2532 (89.7)	291 (10.3)	
Metro status			0.5352
Non-metro	1883 (88.7)	241 (11.4)	
GNO	2140 (89.4)	254 (10.6)	
Other metro	8922 (88.6)	1149 (11.4)	

[1]Age at COVID-19 diagnosis or year 2021

Abbreviation: NHB, non-Hispanic black; NHW, non-Hispanic white, P/M/O, Private/Medicare/other public; GNO, Greater New Orleans

Figure 1. COVID-19 Infection Rate by Race/Ethnicity and COVID-19 Diagnosis Year

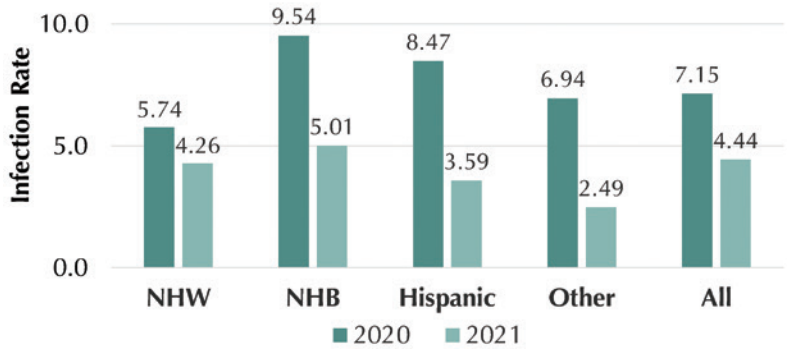


Table 2. Factors Associated with COVID-19 Infection Among Working-Age Women with Precancerous Cervical Lesion Diagnosed in 2009–2021

Variables	OR (95% CI)	aOR (95%CI)	Variables	OR (95% CI)	aOR (95%CI)
Health insurance			Health insurance		
Medicaid vs P/M/O	0.81 (0.61-1.08)	0.73 (0.54-0.98)	Medicaid vs P/M/O	0.81 (0.61-1.08)	0.73 (0.54-0.98)
No/unk vs P/M/O	0.81 (0.68-0.97)	0.80 (0.66-0.97)	No/unk vs P/M/O	0.81 (0.68-0.97)	0.80 (0.66-0.97)
SES			SES		
Group 2 vs 1	0.88 (0.75-1.03)	0.99 (0.84-1.16)	Group 2 vs 1	0.88 (0.75-1.03)	0.99 (0.84-1.16)
Group 3 vs 1	0.79 (0.67-0.93)	0.93 (0.78-1.10)	Group 3 vs 1	0.79 (0.67-0.93)	0.93 (0.78-1.10)
Group 4 vs 1	0.91 (0.78-1.06)	1.10 (0.93-1.30)	Group 4 vs 1	0.91 (0.78-1.06)	1.10 (0.93-1.30)
Group 5 vs 1	0.79 (0.67-0.93)	0.96 (0.80-1.15)	Group 5 vs 1	0.79 (0.67-0.93)	0.96 (0.80-1.15)
Metro status			Metro status		
Non-metro vs GNO	1.08 (0.90-1.30)	1.24 (1.02-1.51)	Non-metro vs GNO	1.08 (0.90-1.30)	1.24 (1.02-1.51)
Other metro vs GNO	1.08 (0.94-1.25)	1.17 (1.01-1.36)	Other metro vs GNO	1.08 (0.94-1.25)	1.17 (1.01-1.36)

[1]Age at COVID-19 diagnosis or year 2021

Abbreviation: aOR, Adjusted OR; NHB, non-Hispanic black; NHW, non-Hispanic white, P/M/O, Private/Medicare/other public; GNO, Greater New Orleans

NAACCR Third-Place Poster

New Hampshire Childhood Cancer Survivor Study: A Qualitative Study

Angela Ricci[a,b]; Meghan Longacre[c]; Nithya Ramesh[c]; Julie Kim[a,b]; Adriane Burke[d]; Judy Rees[e]

Background

- Childhood cancer crude incidence (2013–2017) was reported as being highest in New Hampshire and the Northeast (Siegel 2018).
- In response, the New Hampshire State Legislature set aside funds to explore childhood cancer issues in the Granite State.
- This funding allowed a multidisciplinary team led by the New Hampshire State Cancer Registry (NHSCR) to conduct several childhood cancer projects, including the survivor study, to better understand experiences of childhood cancer patients and their families

Objectives

- Identify core challenges and supports for survivors and parents during active cancer treatment
- Describe survivors' and parents' experiences with the transition to survivorship care
- Elicit recommendations to improve the experience of pediatric cancer and survivorship care

Methods

Study Design

- Qualtrics online survey to collect sociodemographic characteristics
- Focus groups and individual semi-structured interviews to elicit open-ended perspectives
- Project timeline January through June 2021

Recruitment

- Letters mailed to qualified patients and families after review of patient list by oncologist at Dartmouth Hitchcock Medical Center (Table 1)
- Social media advertising (Figure 1)
- Recruitment lasted 2 months (April through May 2021)

Analysis

- Two evaluators from the Center for Program Design and Evaluation (CPDE) conducted focus group (90 minutes) and interviews (60 minutes) following a semi-structured guide
- All sessions conducted virtually over Zoom
- Transcripts were uploaded and coded using Dedoose research data app (version 8.3.47, SocioCultural Research Consultants, LLC) using a mixed deductive-inductive approach
- Coding and thematic statements were developed via consensus between 2 analysts

Challenges

- Short timeline to design and complete the study
  - Issues with use of college or medical center institutional review board (IRB)
  - Challenges with rules regarding participant honorarium
  - IRB-related delays led to very short time left in study period for recruitment and study completion

Table 1. Recruitment Qualifications

Parents of childhood cancer survivors	*Childhood cancer survivors*
≥18 years of age	≤18–25 years of age
English-speaking	English-speaking
Child diagnosed with cancer in the past 10 years and was ≤18 at time of diagnosis	Diagnosed with cancer in the past 10 years
You and your child were New Hampshire residents at the time of treatment; medical treatment may have occurred within or outside New Hampshire (eg, Boston)	You were a New Hampshire resident at the time of treatment; medical treatment may have occurred within or outside of New Hampshire (eg, Boston)
Treatment included either chemotherapy or radiation	Treatment included either chemotherapy or radiation
Your child is currently living and has finished cancer treatment	You have finished cancer treatment

[a]Dartmouth Cancer Center. [b]Children's Hospital at Dartmouth Health. [c]Center for Program Design and Evaluation, Dartmouth College. [d]New Hampshire Department of Health and Human Services. [e]New Hampshire State Cancer Registry.

This content was originally presented as a poster at the NAACCR 2022 Summer Forum, June 14–16, 2022.

Limitations

- Small sample size with only 2 survivor participants
- Parent participants were primarily mothers

Results

Participant characteristics (11 parents, 2 childhood cancer survivors) are shown in Tables 1 and 2. Themes and comments from parents and survivors are noted below.

THEME: Pediatric Cancer is an Isolating Experience

"We kept asking, is there other kids that were [daughter's] age or even older kids, just ***somebody that she could connect with****? I just needed somebody like us." (Parent)*

"It just seemed like a long, lonely trip, even though we had support. It's just the ***most isolating experience*** *I've ever gone through, and it's still to this day, a couple of years later, when I think about how painful it was, I don't wish it on anybody. And nobody can really understand what you're going through, except some other parent who has gone through it because nobody really knows what it's like to watch your child fight to survive." (Parent)*

"I just ***didn't have the energy to be the cancer educator every single time*** *I had a conversation with somebody, so that I found myself isolating from those people. And at a certain point, I found it* ***just to be easier to have no friends****, because then I can just focus on what I was doing." (Parent)*

*"…****as my treatment went on, I lost all of my friends except two****. So that played into effect senior year. I didn't really have anyone. And when I went to school, only two people ever talked to me… people just got caught up in their lives… part of it was they just didn't understand why I still wasn't healthy." (Survivor)*

THEME: Patients and Families Have a Variety of Ongoing Mental Health Needs

About half of participants identified mental health needs including anxiety, depression, and post-traumatic stress disorder for patient, parent, or siblings associated with cancer diagnosis and treatment.

*"****I often felt his quality of life really took a back seat to treating the cancer, which of course, is the most important thing****. I want him alive and healthy first. But a lot of the side effects that came with the emotional piece… and I often felt like I was just yelling into the wind, trying to find some respite for him for the emotional roller coaster that he was going through." (Parent)*

*"****You cannot overstate the effect on the entire family****. This diagnosis didn't just happen to her, it happened to him [her noncancer son], it happened to all of us, and* ***we're still struggling with some of the emotional effects after****…All these pieces come together, the isolation, the fear, the anxiety just seems to linger for a long time after." (Parent)*

THEME: Financial Toxicity is Common (Figure 2)

*"****We could have lost our house****…if we didn't have the support of family that were able to financially help us, because just cutting your income in half for a year and a half, most people can't support that. Even with a good deductible, it's still so expensive, and there has to be something available to families from the State." (Parent, employer-sponsored insurance)*

THEME: Families Experience Lack of Support During Transition to Survivorship

"It felt we were almost going through a war zone for two and a half, three years, and ***then it ended, and we just felt like we were adrift for so long****. Just to have some continued intervention,… whether it's* ***somebody to help me put all those pieces together*** *of the after-effects of both the physical, emotional, and psychological effects after treatment." (Parent)*

"Just to realize that ***even though you're out of treatment not all the services should stop****." (Parent)*

"I wish somebody had told me that, 'After cancer…' I remember talking to somebody and they were like, 'Yea, you'll be just fine.' ***But fine is like, you'll be fine after cancer because you're alive. Right? But what is that reality? A lot of things change****. And if I knew that going into cancer and afterwards, that would have been a bit easier on myself." (Survivor)*

THEME: School Support is Inadequate for Children with Cancer

- Prolonged school interruptions and older children were associated with increased negative impact
- Challenges with continuing accommodations after treatment ended
- Lack of knowledge among school personnel about ongoing or late effects of cancer treatment
- Fatigue with respect to the need to repeatedly advocate for child's needs

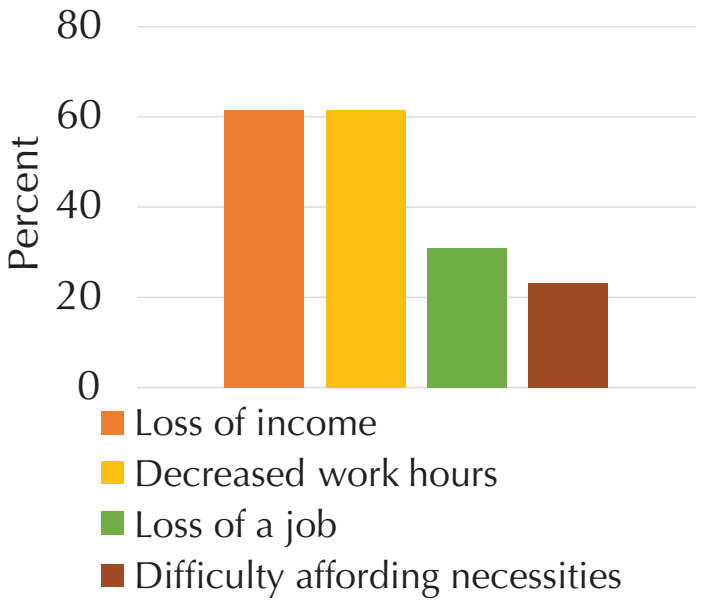
"They [high school teachers] just didn't understand. And I could only explain so much. And there were a couple of teachers that really cared and really tried. But a few of them just, were done. Because I've been out in and out of school for three years. ***So they were like, 'Well you seem fine. I don't understand why you're not able to do the work.'****" (Survivor)*

Figure 1. Social Media Ad

Table 2. Patient Cancer History		
Characteristic	*n*	*%*
Patient cancer history*		
Sex (female)		
Female	9	69.2
Male	4	
Age at diagnosis (y)		
0–4	5	38.5
5–11	3	23.1
≥12	5	38.5
Cancer type		
Leukemia	8	61.5
Lymphoma	2	15.4
Solid tumor	3	23.1
Treatment type		
Chemotherapy	6	46.2
Chemotherapy + surgery	5	38.5
Chemotherapy + radiation + other	2	15.4
Travel time for treatments		
30–60 minutes	7	53.8
1–2 hours	4	30.8
>2 hours	2	15.4
Family-level characteristics*		
Insurance status		
Employer-sponsored	5	38.5
State-sponsored/Medicaid	5	38.5
Both	2	15.4
Household income		
<$50,000	5	38.5
$50,000–$99,999	4	30.8
≥$100,000	4	30.8
Participant demographics^		
Ethnicity/race		
White	10	76.9
Black	2	15.4
Hispanic	1	7.7
Sex		
Female	11	84.6
Male	2	

* Patient and family characteristics at time of diagnosis with cancer.
^ Characteristics of study participant, patient, or parent.

Figure 2. Financial Burdens Associated with Childhood Cancer Among Study Participants



Discussion

Our study identified significant challenges faced by New Hampshire childhood cancer patients and their families both during and after treatment. Particular areas of need include:

- Interpersonal and mental health support
- Financial resources
- School-based support
- More comprehensive and integrated survivorship care
- Opportunities to connect to New Hampshire families experiencing childhood cancer

Future Directions

- Ongoing discussion with New Hampshire Department of Health through Childhood Cancer Projects regarding increasing statewide networking and resources
- Evaluating DH resources for partnership for childhood cancer survivor care at NCCC
- Considering additional grant support for ongoing study and project development in the field of childhood cancer survivor care for residents of Northern New England

Acknowledgements

We thank our study participants—survivors of childhood cancer and their families. We thank the New Hampshire Drinking Water and Groundwater Trust Fund for funding this study.

Journal of Registry Management Continuing Education Quiz—WINTER 2022

THE CASE OF THE MISSING 2020 CANCERS: USING CLAIMS DATA TO INVESTIGATE A DEFICIT IN INCIDENT CANCER CASE REPORTS TO THE NEW YORK STATE CANCER REGISTRY

After reading the educational posters and completing the quiz, participants will be able to:

- Describe how changes in access to medical care due to COVID-19 likely affect cancer incidence rates
- Understand how state-specific patient claims databases can be leveraged to get an early picture of the impact of COVID-19 on cancer rates.

1. The COVID-19 pandemic had little impact on the amount or type of health care utilized in 2020.
   a) True
   b) False
   c) Unknown if true or false.
2. Cancer rates for 2020 are likely to be decreased due to which of the following?
   a) Decreases in cancer risk in the underlying population
   b) Changes in the amount and type of health care accessed in 2020
   c) The reason behind the decrease in rates is unknown
   d) 2020 rates are not expected to decrease
3. The extent of the impact of public health orders in New York State on health care utilization can be assumed to be the same in other states.
   a) True
   b) False
4. The percent change in the overall number of 2020 versus 2019 cancer-related claims in New York State was greatest during which of the following time frames?
   a) January–February 2020
   b) April 2020
   c) July–December 2020
   d) The change was constant for all months of 2020
5. The percent change in the overall number of 2020 versus 2019 cancer-related claims in New York State was greatest for which of the following encounter types?
   a) Ambulatory surgery center claims
   b) Emergency department claims
   c) Inpatient claims
   d) Outpatient claims
6. The percent change in the overall number of 2020 versus 2019 cancer-related claims in New York State was greatest for which of the following age groups?
   a) 0–19 years
   b) 20–29 years
   c) 30–39 years
   d) 40–49 years
   e) 50–79 years
   f) ≥80 years
7. The percent change in the overall number of 2020 versus 2019 cancer-related claims in New York state was greatest for which of the following racial/ethnic groups?
   a) Non-Hispanic White
   b) Non-Hispanic Black
   c) Non-Hispanic Asian/Pacific Islander
   d) Hispanic
8. How does the percent change in the overall number of 2020 versus 2019 cancer-related claims for New York City compare to the rest of New York State?
   a) The same
   b) Higher in New York City
   c) Lower in New York City
   d) Unknown
9. The percent change in cancer-related claims only impacted the early months (March–April) of 2020.
   a) True
   b) False
10. The percent change in cancer-related claims was a decrease regardless of race, ethnicity, age, month of diagnosis, geographic location, or facility type. Does this mean that the impact of COVID-19 on access to medical care affected all patients equally?
    a) True
    b) False

Purchase Quiz to Earn CE:

1. Go to **http://www.cancerregistryeducation.org/jrm-quizzes**
2. Select quiz and "Add to Cart" (You may be prompted to login using your NCRA login).
3. Continue through the checkout process.
4. Once purchase is complete, the quiz will load automatically into "My Learning Activities" page.

Journal of Registry Management
Volume 49, Spring 2022 to Winter 2022

Reviewer acknowledgement: *JRM gratefully acknowledges the individuals who have served as manuscript reviewers or have otherwise assisted in the review process during the past year. Their wise counsel and contributions to the Journal have been most valued.*

Multiple Author Index—Vol. 49 (2022)

A

B

C

D

Dailey, Shantel

Traverso-Ortiz M, Duran D, Mesnard M, Ng D, Dailey S. Results of Year 2 Data Quality Evaluation of CDC's National Program of Cancer Registries: Weighing the Evidence, Identifying Research Gaps, and Evaluating Outputs of a Prevention Research Agenda. Summer;49(2):75-78.

Doherty, Jennifer A.

Lawson-Michod KA, Carter M, Yoder V, McCarty RD, Bateman C, Millar MM, Doherty JA. Improving Precision of Do Not Contact Codes: Results of a Manual Review to Inform Coding and Case Contact Procedures. Winter;49(4):126-131.

Duke, Wes

Duke W, Alverson CJ, Evans SP, Atkinson M, Ailes EC. Using a Birth Defects Surveillance Program to Enhance Existing Surveillance of Stillbirth. Spring;49(1):17-22.

Duran, Denise

Traverso-Ortiz M, Duran D, Mesnard M, Ng D, Dailey S. Results of Year 2 Data Quality Evaluation of CDC's National Program of Cancer Registries: Weighing the Evidence, Identifying Research Gaps, and Evaluating Outputs of a Prevention Research Agenda. Summer;49(2):75-78.

Durbin, Eric B.

Rollison DE, Levin GM, Warner JL, Pinder R, Havener LA, Behera M, Post AR, Gopalakrishnan R, Burbin EB. Current and Emerging Informatics Initiatives Impactful to Cancer Registries. Winter;49(4):153-160.

E

Edwards, Patrick

Edwards P, Bernacet A, Tangka FKL, Pordell P, Beizer J, Wilson R, Blumenthal W, Jones SF, Cole-Beebe M, Subramanian S. Operational Characteristics of Central Cancer Registries That Support the Generation of High-Quality Surveillance Data. Spring;49(1):10-16.

Elayavilli, Ravikumar K.

Miller DM, Shalhout SZ, Saqlain F, Patel VA, Tsai KY, Elayavilli RK, Louv B, Brownell I, Wong MK. The Merkel Cell Carcinoma Patient Registry: From Promise to Prototype to Patient. Spring; 49(1):4-9.

Eliassen, M. Scottie

Riddle BL, Eliassen MS, Celaya MO, Bancroft C, Johnson A, Oh J, Chawla C, Rees JR. Identifying Duplicate Cancer Cases Across State Boundaries: Pseudonymization as a Method to Encrypt Identifiers. Fall;49(3):88-91.

Evans, Shannon P.

Duke W, Alverson CJ, Evans SP, Atkinson M, Ailes EC. Using a Birth Defects Surveillance Program to Enhance Existing Surveillance of Stillbirth. Spring;49(1):17-22.

F

Feliciano, Sarah Moncrief

Kuliszewski MG, Sickler H, Sango-Jordan M, Germain SS, Feliciano SM, Vyce N, Schymura MJ. Participation Rates and Characteristics of Participants and Nonparticipants in Cancer Patient Contact Studies in New York State. Winter;49(4):194-195.

Firth, Albert

Ostrom Q, Kruchko C, Neff C, Firth A, Sherman R. The Central Brain Tumor Registry of the United States Histopathological Grouping Scheme Provides Clinically Relevant Brain and Other Central Nervous System Categories for Cancer Registry Data. Winter;49(4):139-152.

Fritzler, Jessica

Fritzler J, Spivak G. Mapping Cancer Incidence among Hereditary Cancer Types in Michigan, 2014–2018. Summer;49(2):69-70.

Fuller, Heather

McNellis G, Fuller H. Demystifying the RCRS Guidelines for Breast BCSRT Measure. Summer;49(2):56-57.

G

Ganz, Connie

Wu Q, Ganz C, Li L. Data Quality Control for Electronic Pathology Reporting. Fall;49(3):95-96.

Germain, Stephanie Saint

Kuliszewski MG, Sickler H, Sango-Jordan M, Germain SS, Feliciano SM, Vyce N, Schymura MJ. Participation Rates and Characteristics of Participants and Nonparticipants in Cancer Patient Contact Studies in New York State. Winter;49(4):194-195.

Gopalakrishnan, Rajan

Rollison DE, Levin GM, Warner JL, Pinder R, Havener LA, Behera M, Post AR, Gopalakrishnan R, Burbin EB. Current and Emerging Informatics Initiatives Impactful to Cancer Registries. Winter;49(4):153-160.

Graber, Judith M.

Stroup AM, Harris G, Wilson M, Black TM, Graber JM. New Jersey's Response to the Call for Firefighter Cancer Data: Creating a Retrospective Cohort to Study Cancer Mortality among New Jersey Firefighters through Data Linkages and an Honest Broker. Spring;49(1):39-40.

H

Haggan, Kim

Bancroft C, Haggan K, Boris K, Cross J, Huston S, Yob D. Responding to Constituent Cancer Concerns: The Maine Cancer Registry's Cancer Inquiry Group. Fall;49(3):92-93.

J

K

L

Lawson-Michod, Katherine A.

Lawson-Michod KA, Carter M, Yoder V, McCarty RD, Bateman C, Millar MM, Doherty JA. Improving Precision of Do Not Contact Codes: Results of a Manual Review to Inform Coding and Case Contact Procedures. Winter;49(4):126-131.

Lefante, Tina

Hsieh M, Lefante T, Yi Y, Wu X. Racial/Ethnic Disparities in COVID-19 Infection Among Working-Age Women with Precancerous Cervical Lesion. Winter;49(4):196-197.

Lemons, Debbi

Zimmerman H, Lemons D. The Statewide Voter File Proves Useful to Obtain Physical Address for Cases with Only PO Box Reported. Spring;49(1):37-38.

Levin, Gary M.

Rollison DE, Levin GM, Warner JL, Pinder R, Havener LA, Behera M, Post AR, Gopalakrishnan R, Burbin EB. Current and Emerging Informatics Initiatives Impactful to Cancer Registries. Winter;49(4):153-160.

Hernandez MN, Levin GM, Hylton T, Pordell P, White A, Senkomago V. Integration of Cancer Screening Data into Routine Cancer Surveillance Systems: A Florida Pilot Project. Winter;49(4):132-138.

Li, Lifeng

Wu Q, Ganz C, Li L. Data Quality Control for Electronic Pathology Reporting. Fall;49(3):95-96.

Longacre, Meghan

Ricci A, Longacre M, Ramesh N, Kim J, Burke A, Rees J. New Hampshire Childhood Cancer Survivor Study: A Qualitative Study. Winter;49(4):198-200.

Louv, Bill

Miller DM, Shalhout SZ, Saqlain F, Patel VA, Tsai KY, Elayavilli RK, Louv B, Brownell I, Wong MK. The Merkel Cell Carcinoma Patient Registry: From Promise to Prototype to Patient. Spring; 49(1):4-9.

Luna, Guillermo T.

Ortiz MA, Cintrón CRT, Ruiz YR, Zegarra DEZ, Luna GT. The Puerto Rico Central Cancer Registry—Geocoding Project: Right on Track. Spring;49(1):41-42.

M

MacKinnon, Jill A.

Boone CR, MacKinnon JA. South Carolina Innovative Use of FLccSC with Contractors. Summer;49(2):79-80.

Mariotto, Angela B.

Weir HK, Bryant H, Turner D, Coleman MP, Mariotto AB, Spika D, Matz M, Harewood R, Tucker TC, Allemani C; Canada–US CONCORD Working Group. Population-Based Cancer Survival in Canada and the United States by Socio-economic status: Findings From the CONCORD-2 Study. Spring;49(1):23-33.

Tatalovich Z, Stinchcomb DG, Mariotto A, Ng D, Stevens JL, Coyle LM, Penberthy L. Assessment of Interstate Residential Mobility of SEER Patients: SEER and LexisNexis Residential Address Linkage. Winter;49(4):109-113.

Morawski BM, Hsieh M, Wu M, Sherman R, Mariotto AB, Johnson CJ. Determining Fitness for Use of SEER Cause-Specific Cause of Death in Analyses of Cause-Specific Survival. Winter;49(4):177-189.

Matz, Melissa

Weir HK, Bryant H, Turner D, Coleman MP, Mariotto AB, Spika D, Matz M, Harewood R, Tucker TC, Allemani C; Canada–US CONCORD Working Group. Population-Based Cancer Survival in Canada and the United States by Socio-economic status: Findings From the CONCORD-2 Study. Spring;49(1):23-33.

McCarty, Rachel D.

Lawson-Michod KA, Carter M, Yoder V, McCarty RD, Bateman C, Millar MM, Doherty JA. Improving Precision of Do Not Contact Codes: Results of a Manual Review to Inform Coding and Case Contact Procedures. Winter;49(4):126-131.

McCoy, Markie P.

McCoy MP. Expanding Research Access to West Virginia Cancer Registry Data. Spring; 49(1):45.

McNellis, Gina

McNellis G, Fuller H. Demystifying the RCRS Guidelines for Breast BCSRT Measure. Summer;49(2):56-57.

Mesnard, Mary

Traverso-Ortiz M, Duran D, Mesnard M, Ng D, Dailey S. Results of Year 2 Data Quality Evaluation of CDC's National Program of Cancer Registries: Weighing the Evidence, Identifying Research Gaps, and Evaluating Outputs of a Prevention Research Agenda. Summer;49(2):75-78.

Millar, Morgan M.

Lawson-Michod KA, Carter M, Yoder V, McCarty RD, Bateman C, Millar MM, Doherty JA. Improving Precision of Do Not Contact Codes: Results of a Manual Review to Inform Coding and Case Contact Procedures. Winter;49(4):126-131.

Miller, David M.

Miller DM, Shalhout SZ, Saqlain F, Patel VA, Tsai KY, Elayavilli RK, Louv B, Brownell I, Wong MK. The Merkel Cell Carcinoma Patient Registry: From Promise to Prototype to Patient. Spring; 49(1):4-9.

Q

Qiao, Baozhen

Qiao B, Austin AA, Schymura MJ, Zhang X, Sherman CG. Identifying Factors Associated with Loss to Follow-up Among Patients Reported to the New York State Cancer Registry. Winter;49(4):161-169.

R

Ramesh, Nithya

Ricci A, Longacre M, Ramesh N, Kim J, Burke A, Rees J. New Hampshire Childhood Cancer Survivor Study: A Qualitative Study. Winter;49(4):198-200.

Reed, Rani

Ogushi A, Reed R. Nevada Central Cancer Registry's Legislative Accomplishments. Fall;49(3):97-98.

Rees, Judy R.

Ricci A, Longacre M, Ramesh N, Kim J, Burke A, Rees J. New Hampshire Childhood Cancer Survivor Study: A Qualitative Study. Winter;49(4):198-200.

Riddle BL, Eliassen MS, Celaya MO, Bancroft C, Johnson A, Oh J, Chawla C, Rees JR. Identifying Duplicate Cancer Cases Across State Boundaries: Pseudonymization as a Method to Encrypt Identifiers. Fall;49(3):88-91.

Ricci, Angela

Ricci A, Longacre M, Ramesh N, Kim J, Burke A, Rees J. New Hampshire Childhood Cancer Survivor Study: A Qualitative Study. Winter;49(4):198-200.

Riddle, Bruce L.

Riddle BL, Eliassen MS, Celaya MO, Bancroft C, Johnson A, Oh J, Chawla C, Rees JR. Identifying Duplicate Cancer Cases Across State Boundaries: Pseudonymization as a Method to Encrypt Identifiers. Fall;49(3):88-91.

Rollison, Dana E.

Rollison DE, Levin GM, Warner JL, Pinder R, Havener LA, Behera M, Post AR, Gopalakrishnan R, Burbin EB. Current and Emerging Informatics Initiatives Impactful to Cancer Registries. Winter;49(4):153-160.

Ruiz, Yadira R.

Ortiz MA, Cintrón CRT, Ruiz YR, Zegarra DEZ, Luna GT. The Puerto Rico Central Cancer Registry—Geocoding Project: Right on Track. Spring;49(1):41-42.

S

Sango-Jordan, Marilyn

Kuliszewski MG, Sickler H, Sango-Jordan M, Germain SS, Feliciano SM, Vyce N, Schymura MJ. Participation Rates and Characteristics of Participants and Nonparticipants in Cancer Patient Contact Studies in New York State. Winter;49(4):194-195.

Santiago, Patti M.

Cheteri MK, Shridhar N, Crain P, Peterson J, Santiago PM. Reducing Morbidity and Mortality for Washington State Residents Impacted by Heritable Cancer Syndromes (HBOC and LS) by Enhancing Public Health Surveillance and Increasing Appropriate Referrals for Genetic Services. Spring;49(1):43-44.

Saqlain, Farees

Miller DM, Shalhout SZ, Saqlain F, Patel VA, Tsai KY, Elayavilli RK, Louv B, Brownell I, Wong MK. The Merkel Cell Carcinoma Patient Registry: From Promise to Prototype to Patient. Spring;49(1):4-9.

Schymura, Maria J.

Qiao B, Austin AA, Schymura MJ, Zhang X, Sherman CG. Identifying Factors Associated with Loss to Follow-up Among Patients Reported to the New York State Cancer Registry. Winter;49(4):161-169.

Kuliszewski MG, Sickler H, Sango-Jordan M, Germain SS, Feliciano SM, Vyce N, Schymura MJ. Participation Rates and Characteristics of Participants and Nonparticipants in Cancer Patient Contact Studies in New York State. Winter;49(4):194-195.

Zhang X, Kuliszewski MG, Kahn AR, Schymura MJ. Early COVID-19 Hospitalizations Among New York State Residents with a History of Invasive Cancer. Winter;49(4):114-125.

Zhang X, Kuliszewski MG, Kahn AR, Schymura MJ. The Case of the Missing 2020 Cancers: Using Claims Data to Investigate a Deficit in Incident Cancer Case Reports to the New York State Cancer Registry in 2020. Winter;49(4):170-176.

Senkomago, Virginia

Hernandez MN, Levin GM, Hylton T, Pordell P, White A, Senkomago V. Integration of Cancer Screening Data into Routine Cancer Surveillance Systems: A Florida Pilot Project. Winter;49(4):132-138.

Shalhout, Sophia Z.

Miller DM, Shalhout SZ, Saqlain F, Patel VA, Tsai KY, Elayavilli RK, Louv B, Brownell I, Wong MK. The Merkel Cell Carcinoma Patient Registry: From Promise to Prototype to Patient. Spring;49(1):4-9.

Sheridan, Cheryl

Sheridan C. COVID-19 Response to Cancer Conference. Summer;49(2):62-63.

Sherman, Colleen G.

Qiao B, Austin AA, Schymura MJ, Zhang X, Sherman CG. Identifying Factors Associated with Loss to Follow-up Among Patients Reported to the New York State Cancer Registry. Winter;49(4):161-169.

Sherman, Recinda

Ostrom Q, Kruchko C, Neff C, Firth A, Sherman R. The Central Brain Tumor Registry of the United States Histopathological Grouping Scheme Provides Clinically Relevant Brain and Other Central Nervous System Categories for Cancer Registry Data. Winter;49(4):139-152.

T

V

Key Word Index—Vol. 49 (2022)

D

E

F

S

Social Determinants

Tatalovich Z, Stinchcomb DG, Mariotto A, Ng D, Stevens JL, Coyle LM, Penberthy L. Assessment of Interstate Residential Mobility of SEER Patients: SEER and LexisNexis Residential Address Linkage. Winter;49(4):109-113.

Stillbirth

Duke W, Alverson CJ, Evans SP, Atkinson M, Ailes EC. Using a Birth Defects Surveillance Program to Enhance Existing Surveillance of Stillbirth. Spring;49(1):17-22.

Surveillance

Duke W, Alverson CJ, Evans SP, Atkinson M, Ailes EC. Using a Birth Defects Surveillance Program to Enhance Existing Surveillance of Stillbirth. Spring;49(1):17-22.

Surveillance, Epidemiology, And End Results (Seer) Program

Lawson-Michod KA, Carter M, Yoder V, McCarty RD, Bateman C, Millar MM, Doherty JA. Improving Precision of Do Not Contact Codes: Results of a Manual Review to Inform Coding and Case Contact Procedures. Winter;49(4):126-131.

Tatalovich Z, Stinchcomb DG, Mariotto A, Ng D, Stevens JL, Coyle LM, Penberthy L. Assessment of Interstate Residential Mobility of SEER Patients: SEER and LexisNexis Residential Address Linkage. Winter;49(4):109-113.

Survival

Weir HK, Bryant H, Turner D, Coleman MP, Mariotto AB, Spika D, Matz M, Harewood R, Tucker TC, Allemani C; Canada–US CONCORD Working Group. Population-Based Cancer Survival in Canada and the United States by Socio-economic status: Findings From the CONCORD-2 Study. Spring;49(1):23-33.

Qiao B, Austin AA, Schymura MJ, Zhang X, Sherman CG. Identifying Factors Associated with Loss to Follow-up Among Patients Reported to the New York State Cancer Registry. Winter;49(4):161-169.

Survival Estimates

Morawski BM, Hsieh M, Wu M, Sherman R, Mariotto AB, Johnson CJ. Determining Fitness for Use of SEER Cause-Specific Cause of Death in Analyses of Cause-Specific Survival. Winter;49(4):177-189.

T

Tumor Registry

Miller DM, Shalhout SZ, Saqlain F, Patel VA, Tsai KY, Elayavilli RK, Louv B, Brownell I, Wong MK. The Merkel Cell Carcinoma Patient Registry: From Promise to Prototype to Patient. Spring; 49(1):4-9.

U

United States

Weir HK, Bryant H, Turner D, Coleman MP, Mariotto AB, Spika D, Matz M, Harewood R, Tucker TC, Allemani C; Canada–US CONCORD Working Group. Population-Based Cancer Survival in Canada and the United States by Socio-economic status: Findings From the CONCORD-2 Study. Spring;49(1):23-33.

W

World Health Organization Classification Of Tumors Of The Central Nervous System

Ostrom Q, Kruchko C, Neff C, Firth A, Sherman R. The Central Brain Tumor Registry of the United States Histopathological Grouping Scheme Provides Clinically Relevant Brain and Other Central Nervous System Categories for Cancer Registry Data. Winter;49(4):139-152.

Title Index—Vol. 49 (2022)

L

Learning from our Children: Adapting Data Management in an Adult World

Costantini A, Adams C. Learning from our Children: Adapting Data Management in an Adult World. Summer;49(2):71-74.

Linking the All-Payers Claims Database (APCD) to the Colorado Central Cancer Registry: An Evaluation

Arend J. Linking the All-Payers Claims Database (APCD) to the Colorado Central Cancer Registry: An Evaluation. Spring;49(1):35-36.

M

Mapping Cancer Incidence among Hereditary Cancer Types in Michigan, 2014–2018

Fritzler J, Spivak G. Mapping Cancer Incidence among Hereditary Cancer Types in Michigan, 2014–2018. Summer;49(2):69-70.

N

Nevada Central Cancer Registry's Legislative Accomplishments

Ogushi A, Reed R. Nevada Central Cancer Registry's Legislative Accomplishments. Fall;49(3):97-98.

New Hampshire Childhood Cancer Survivor Study: A Qualitative Study

Ricci A, Longacre M, Ramesh N, Kim J, Burke A, Rees J. New Hampshire Childhood Cancer Survivor Study: A Qualitative Study. Winter;49(4):198-200.

New Jersey's Response to the Call for Firefighter Cancer Data: Creating a Retrospective Cohort to Study Cancer Mortality among New Jersey Firefighters through Data Linkages and an Honest Broker

Stroup AM, Harris G, Wilson M, Black TM, Graber JM. New Jersey's Response to the Call for Firefighter Cancer Data: Creating a Retrospective Cohort to Study Cancer Mortality among New Jersey Firefighters through Data Linkages and an Honest Broker. Spring;49(1):39-40.

O

Operational Characteristics of Central Cancer Registries that Support the Generation of High-Quality Surveillance Data

Edwards P, Bernacet A, Tangka FKL, Pordell P, Beizer J, Wilson R, Blumenthal W, Jones SF, Cole-Beebe M, Subramanian S. Operational Characteristics of Central Cancer Registries That Support the Generation of High-Quality Surveillance Data. Spring;49(1):10-16.

P

Participation Rates and Characteristics of Participants and Nonparticipants in Cancer Patient Contact Studies in New York State

Kuliszewski MG, Sickler H, Sango-Jordan M, Germain SS, Feliciano SM, Vyce N, Schymura MJ. Participation Rates and Characteristics of Participants and Nonparticipants in Cancer Patient Contact Studies in New York State. Winter;49(4):194-195.

Population-Based Cancer Survival in Canada and the United States by Socio-Economic Status: Findings from the CONCORD-2 Study

Weir HK, Bryant H, Turner D, Coleman MP, Mariotto AB, Spika D, Matz M, Harewood R, Tucker TC, Allemani C; Canada–US CONCORD Working Group. Population-Based Cancer Survival in Canada and the United States by Socio-economic status: Findings From the CONCORD-2 Study. Spring;49(1):23-33.

Prostate Cancer Trends in Montana

Zimmerman HJN. Prostate Cancer Trends in Montana. Winter;49(4):190-193.

R

Racial/Ethnic Disparities in COVID-19 Infection Among Working-Age Women with Precancerous Cervical Lesion

Hsieh M, Lefante T, Yi Y, Wu X. Racial/Ethnic Disparities in COVID-19 Infection Among Working-Age Women with Precancerous Cervical Lesion. Winter;49(4):196-197.

Reducing Morbidity and Mortality for Washington State Residents Impacted by Heritable Cancer Syndromes (HBOC and LS) by Enhancing Public Health Surveillance and Increasing Appropriate Referrals for Genetic Services

Cheteri MK, Shridhar N, Crain P, Peterson J, Santiago PM. Reducing Morbidity and Mortality for Washington State Residents Impacted by Heritable Cancer Syndromes (HBOC and LS) by Enhancing Public Health Surveillance and Increasing Appropriate Referrals for Genetic Services. Spring;49(1):43-44.

Responding to Constituent Cancer Concerns: The Maine Cancer Registry's Cancer Inquiry Group

Bancroft C, Haggan K, Boris K, Cross J, Huston S, Yob D. Responding to Constituent Cancer Concerns: The Maine Cancer Registry's Cancer Inquiry Group. Fall;49(3):92-93.

Results of Year 2 Data Quality Evaluation of CDC's National Program of Cancer Registries: Weighing the Evidence, Identifying Research Gaps, and Evaluating Outputs of a Prevention Research Agenda

Traverso-Ortiz M, Duran D, Mesnard M, Ng D, Dailey S. Results of Year 2 Data Quality Evaluation of CDC's National Program of Cancer Registries: Weighing the Evidence, Identifying Research Gaps, and Evaluating Outputs of a Prevention Research Agenda. Summer;49(2):75-78.

S

South Carolina Innovative Use of FLccSC with Contractors

Boone CR, MacKinnon JA. South Carolina Innovative Use of FLccSC with Contractors. Summer;49(2):79-80.

Spring 2022 Continuing Education Quiz

Mitchell S, Vida C. Spring 2022 Continuing Education Quiz. Spring;49(1):46.

Summer 2022 Continuing Education Quiz

Mitchell S, Vida C. Summer 2022 Continuing Education Quiz. Summer;49(2):81.

T

The Case of the Missing 2020 Cancers: Using Claims Data to Investigate a Deficit in Incident Cancer Case Reports to the New York State Cancer Registry in 2020

Zhang X, Kuliszewski MG, Kahn AR, Schymura MJ. The Case of the Missing 2020 Cancers: Using Claims Data to Investigate a Deficit in Incident Cancer Case Reports to the New York State Cancer Registry in 2020. Winter;49(4):170-176.

The Central Brain Tumor Registry of the United States Histopathological Grouping Scheme Provides Clinically Relevant Brain and Other Central Nervous System Categories for Cancer Registry Data

Ostrom Q, Kruchko C, Neff C, Firth A, Sherman R. The Central Brain Tumor Registry of the United States Histopathological Grouping Scheme Provides Clinically Relevant Brain and Other Central Nervous System Categories for Cancer Registry Data. Winter;49(4):139-152.

The Merkel Cell Carcinoma Patient Registry: From Promise to Prototype to Patient

Miller DM, Shalhout SZ, Saqlain F, Patel VA, Tsai KY, Elayavilli RK, Louv B, Brownell I, Wong MK. The Merkel Cell Carcinoma Patient Registry: From Promise to Prototype to Patient. Spring; 49(1):4-9.

The Puerto Rico Central Cancer Registry—Geocoding Project: Right on Track

Ortiz MA, Cintrón CRT, Ruiz YR, Zegarra DEZ, Luna GT. The Puerto Rico Central Cancer Registry—Geocoding Project: Right on Track. Spring;49(1):41-42.

The Statewide Voter File Proves Useful to Obtain Physical Address for Cases with Only PO Box Reported

Zimmerman H, Lemons D. The Statewide Voter File Proves Useful to Obtain Physical Address for Cases with Only PO Box Reported. Spring;49(1):37-38.

U

Using a Birth Defects Surveillance Program to Enhance Existing Surveillance of Stillbirth

Duke W, Alverson CJ, Evans SP, Atkinson M, Ailes EC. Using a Birth Defects Surveillance Program to Enhance Existing Surveillance of Stillbirth. Spring;49(1):17-22.

W

Winter 2022 Continuing Education Quiz

Mitchell S, Vida C. Winter 2022 Continuing Education Quiz. Winter;49(4):201.

National Cancer Registrars Association
CALL FOR PAPERS

The *Journal of Registry Management*, official journal of the National Cancer Registrars Association (NCRA), announces a call for original manuscripts on registry methodology or research findings related to the 7 subjects listed below and related topics.

Topics:

1. **Birth Defects Registries**
2. **Cancer Registries**
   a. **AJCC TNM Stage**
   b. **Cancer and Socioeconomic Status**
   c. **Cancer and Health Disparities**
3. **Trauma Registries**
4. **Recruitment, Training, and Retention**
5. **Public Relations**
6. **Quality Review**
7. **Registry Management**

Contributed manuscripts are peer-reviewed prior to publication. Manuscripts of the following types may be submitted for publication:

1. **Methodology Articles** addressing topics of broad interest and appeal to the readership, including methodological aspects of registry organization and operation.
2. **Research articles** reporting findings of original, reviewed, data-based research.
3. **Primers** providing basic and comprehensive tutorials on relevant subjects.
4. **"How I Do It" Articles** describe tips, techniques, or procedures for an aspect of registry operations that the author does particularly well. The "*How I Do It*" feature in the *Journal* provides registrars with an informal forum for sharing strategies with colleagues in all types of registries.
5. **Opinion papers/editorials** including position papers, commentaries, essays, and interviews that analyze current or controversial issues and provide creative, reflective treatments of topics related to registry management.
6. **Bibliographies** which are specifically targeted and of significant interest will be considered.
7. **Letters to the Editor** are also invited.

Address all manuscripts to: Neisha Johnson, *Journal of Registry Management,*
(703) 299-6640 ext. 319, JRMEditor@ncra-usa.org.

Manuscript submission requirements are given in "Information for Authors" found near the back of each *Journal* and on the NCRA website at **http://www.ncra-usa.org/jrm**.

INFORMATION FOR AUTHORS

Journal of Registry Management (JRM), the official journal of the National Cancer Registrars Association, invites submission of original manuscripts on topics related to management of disease registries and the collection, management, and use of cancer, trauma, AIDS, and other disease registry data. Reprinting of previously published material will be considered for publication only when it is of special and immediate interest to the readership. *JRM* encourages authorship by Certified Tumor Registrars (CTRs); special value is placed on manuscripts with CTR collaboration and publication of articles or texts related to the registry profession. CTR continuing education (CE) credits are awarded; a published chapter or full textbook article equals 5 CE hours. Other published articles or documents equal CE hours. All correspondence and manuscripts should be addressed to Neisha Johnson at JRMEditor@ncra-usa.org or (703) 299-6640 ext. 319.

Manuscripts may be submitted for publication in the following categories: **Articles** addressing topics of broad interest and appeal to the readership, including **Methodology papers** about registry organization and operation; **Research papers** reporting findings of original, reviewed, data-based research; **Primers** providing tutorials on relevant subjects; and **"How I Do It"** papers are also solicited. **Opinion papers/editorials** including position papers, commentaries, and essays that analyze current or controversial issues and provide creative, reflective treatments of topics related to registry management; **Letters to the Editor**; and specifically-targeted **Bibliographies** of significant interest are invited.

The following guidelines are provided to assist prospective authors in preparing manuscripts for the *Journal*, and to facilitate technical processing of submissions. Failure to follow the guidelines may delay consideration of your manuscript. Authors who are unfamiliar with preparation and submission of manuscripts for publication are encouraged to contact the Editor for clarification or additional assistance.

Submission Requirements

Manuscripts. The terms manuscripts, articles, and papers are used synonymously herein. Email only submission of manuscripts is encouraged. If not feasible, submit the original manuscript and 4 copies to the Editor. Manuscripts should be double-spaced on white 8-1/2" x 11" paper, with margins of at least 1 inch. Use only letter-quality printers; poor quality copies will not be considered. Number the manuscript pages consecutively with the (first) title page as page one, followed by the abstract, text, references, and visuals. The accompanying cover letter should include the name, mailing address, email address, and telephone number of the corresponding author. For electronic submission, files should be IBM-compatible format in Corel WordPerfect™, Microsoft® Word for Windows®, or converted to ASCII code.

Manuscripts (Research Articles). Articles should follow the standard format for research reporting (Introduction, Methods, Results, Discussion, References), and the submission instructions outlined above. The introduction will normally include background information, and a rationale/justification as to why the subject matter is of interest. The discussion often includes a conclusion subsection. Comprehensive references are encouraged, as are an appropriate combination of tables and figures (graphs).

Manuscripts (Methodology/Process Papers). Methodology papers should follow the standard format for research reporting (Introduction, Methods, Results, Discussion), or for explanatory papers not reporting results (Introduction, Methods, Discussion), as well as the submission instructions outlined above.

Manuscripts ("How I Do It" articles). The *"How I Do It"* feature in the *Journal* provides registrars with a forum for sharing strategies with colleagues in all types of registries. These articles describe tips, techniques, or procedures for an aspect of registry operations that the author does particularly well. When shared, these innovations can help registry professionals improve their skills, enhance registry operations, or increase efficiency.

"How I Do It" articles should be 1,500 words or less (excepting references) and can contain up to 2 tables or figures. To the extent possible, the standard headings (Introduction, Methods, Results, Discussion) should be used. If results are not presented, that section may be omitted. Authors should describe the problem or issue, their solution, advantages (and disadvantages) to the suggested approach, and their conclusion. All submitted *"How I Do It"* articles will have the benefit of peer/editorial review.

Authors. Each author's name, degrees, certifications, title, professional affiliation, and email address must be noted on the title page exactly as it is to appear in publication. The corresponding author should be noted, with mailing address included. Joint authors should be listed in the order of their contribution to the work. Generally, a maximum of 6 authors for each article will be listed.

Title. Authors are urged to choose a title that accurately and concisely describes the content of the manuscript. Every effort will be made to use the title as submitted; however, *Journal of Registry Management* reserves the right to select a title that is consistent with editorial and production requirements.

Abstract. A brief abstract must accompany each article or research paper. The abstract should summarize the main point(s) and quickly give the reader an understanding of the manuscript's content. It should be placed on a page by itself, immediately following the title page.

Length. Authors are invited to contact the Editor regarding submission of markedly longer manuscripts.

Style. Prepare manuscripts using the *American Medical Association Manual of Style,* 11th ed. (2020).

Visuals. Use visuals selectively to supplement the text. Visual elements—charts, graphs, tables, diagrams, and figures—will be reproduced exactly as received. Copies must be clear and properly identified, and preferably emailed. Each visual must have a brief, self-explanatory title. Submit each visual on a separately numbered page at the end of the manuscript, following the references.

Attribution. Authors are to provide appropriate acknowledgment of products, activities, and support especially for those articles based on, or utilizing, registry data (including acknowledgment of hospital and central registrars). Appropriate attribution is also to be provided to acknowledge federal funding sources of registries from which the data are obtained.

References. References should be carefully selected, and relevant. References must be numbered in order of their appearance in the text. At the end of the manuscript, list the references as they are cited; do not list references alphabetically. Journal citations should include author, title, journal, year, volume, issue, and pages. Book citations should include author, title, city, publisher, year, and pages. *Authors are responsible for the accuracy of all references.* Examples:

1. LeMaster PL, Connell CM. Health education interventions among Native Americans: a review and analysis. *Health Education Quarterly.* 1995;21(4):521-538.

2. Hanks GE, Myers CE, Scardino PT. Cancer of the prostate. In: DeVita VT, Hellman S, Rosenberg SA. *Cancer: Principles and Practice of Oncology*. 4th ed. Philadelphia, PA: J.B. Lippincott Co.; 1993:1073–1113.

Key words. Authors are requested to provide up to 5, alphabetized key words or phrases which will be used in compiling the Annual Subject Index.

Affirmations

Copyright. Authors submitting a manuscript do so on the understanding that if it is accepted for publication, copyright in the article, including the right to reproduce the article in all forms and media, shall be assigned exclusively to NCRA. NCRA will not refuse any reasonable requests by the author(s) for permission to reproduce any of his or her contributions to the *Journal*. Further, the manuscript's accompanying cover letter, signed by all authors, must include the following statement: "We, the undersigned, transfer to the National Cancer Registrars Association, the copyright for this manuscript in the event that it is published in *Journal of Registry Management.*" Failure to provide the statement will delay consideration of the manuscript. It is the author's responsibility to obtain necessary permission when using material (including graphs, charts, pictures, etc) that has appeared in other published works.

Originality. Articles are reviewed for publication assuming that they have not been accepted or published previously and are not under simultaneous consideration for publication elsewhere. If the article has been previously published or significantly distributed, this should be noted in the submission for consideration.

Editing

Journal of Registry Management reserves the right to edit all contributions for clarity and length. Minor changes (punctuation, spelling, grammar, syntax) will be made at the discretion of the editorial staff. Substantive changes will be verified with the author(s) prior to publication.

Peer Review

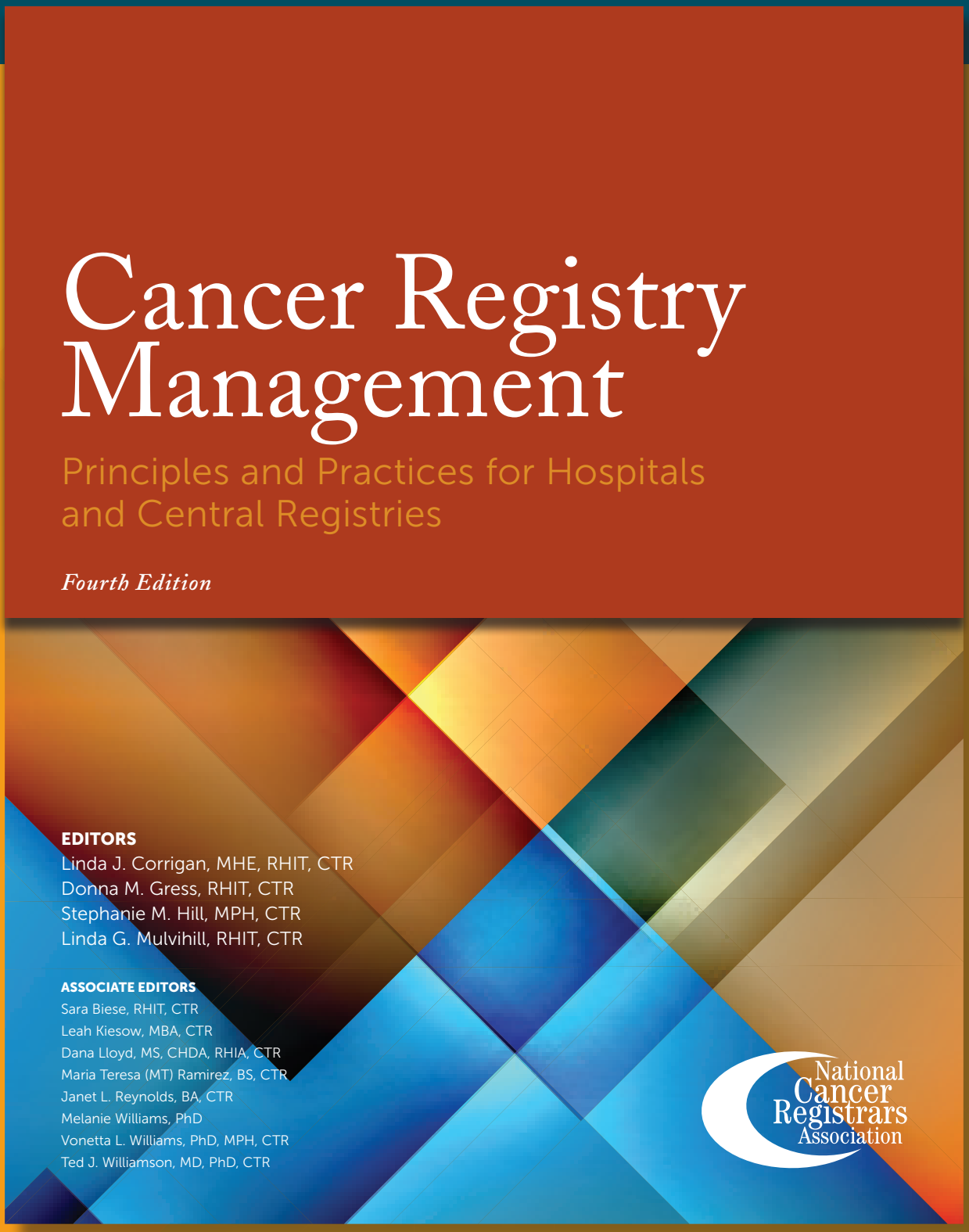
Contributed manuscripts are peer-reviewed prior to publication, generally by 3 reviewers. The *Journal* Editor makes the final decision regarding acceptance of manuscripts. Receipt of manuscripts will be acknowledged promptly, and corresponding authors will be advised of the status of their submission as soon as possible.

Reprints

Authors receive 5 complimentary copies of the *Journal* in which their manuscript appears. Additional copies of reprints may be purchased from the NCRA Executive Office.

A definitive update of the premier textbook and desk reference for the cancer registry profession.

Order print or e-book at
www.ncra-usa.org/4thEdition *today!*



Journal of Registry Management

NCRA Executive Office
1330 Braddock Place
#520
Alexandria, VA 22314

PRESORTED
STANDARD
U.S. POSTAGE
PAID
DULLES, VA
PERMIT #6418

ADDRESS SERVICE REQUESTED

SUSTAINABLE FORESTRY INITIATIVE
Certified Sourcing
www.sfiprogram.org
SFI-00359

100% green power wind & solar

Printed on SFI fiber sourcing paper with non-petroleum, vegetable based inks and manufactured with renewable electricity