ISSN: 1945-6123

National Cancer Registrars Association

Journal of Registry Management

Fall 2021 | Volume 48 | Number 3

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

National Cancer Registrars Association

FINDING YOUR DREAM JOB IS JUST A CLICK AWAY

Visit NCRA's Job Bank at **www.ncra-usa.org/job**!

VIEW JOBS
Access the newest jobs available to cancer registrars seeking employment.

POST AN ANONYMOUS RESUME
Whether you are actively seeking work or curious about what is available, your online resume is your ticket to great job offers!

DESIGN A PERSONAL JOB ALERT
Create job alerts and never let a matching job opportunity pass you by! New jobs that match your search criteria will be e-mailed directly to you.

CREATE A FREE JOB SEEKER ACCOUNT
Your personal job seeker account will allow you to find jobs, manage your resume, and set-up job alerts. Get started today!

NEED A CLINICAL PLACEMENT SITE?
NCRA's Job Bank also includes sites where students can complete their clinical hours and search for mentors.

NCRA'S JOB BANK www.ncra-usa.org/job

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:
Danette Clark, BS, RMA, AAS, CTR, Editor-in-Chief
Email: JRMEditor@ncra-usa.org
Telephone: (973) 971-5189

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

Fall 2021 • Volume 48 • Number 3

Contents

Letters from the Editors
Danette Clark BS, RMA, AAS, CTR; Recinda Sherman, PhD, CTR 87

Original Articles
Challenges of Medical Record Abstraction in a Long-Term Follow-up Study 88
Muhammad F. Ahmed, MBBS, MPH; Andrea Galfo, MPH; Wendy Huggins, BS; Lisa E. Paddock, PhD, MPH; Antoinette M. Stroup, PhD; Jyoti Malhotra, MD, MPH

Association between Cancer Risk and Behaviors Adherent to Cancer Prevention Recommendations in Ontario, Canada 92
Ying Wang, MSc; Mohammad Haque, MPH; Stephanie Young, MPH; Michelle Cotterchio, PhD; Rebecca Truscott, MHSc, RD

Cancer Among Refugees Resettled to Idaho During 2008–2019: A Proof-of-Concept Study 104
Bożena M. Morawski, PhD, MPH; Randi Rycroft, MSPH, CTR; Christopher J. Johnson, MPH

Bias Introduced by Relying on Incomplete Electronic Pathology Reporting for Rapid Case Ascertainment in Patient Contact Studies 110
Margaret Gates Kuliszewski, ScD; Jovanka N. Harrison, PhD; Maria J. Schymura, PhD

Predictors of Survival among Male and Female Patients with Malignant Pleural Mesothelioma: A Random Survival Forest Analysis of Data from the 2000–2017 Surveillance, Epidemiology, and End Results Program 118
Bian Liu, PhD; Li Niu, PhD; Francis Boscoe, PhD; Furrina F. Lee, PhD

Health Care Utilization Prior to Ovarian Cancer Diagnosis in Publicly Insured Individuals in New York State 126
Margaret Gates Kuliszewski, ScD; Francis P. Boscoe, PhD; Victoria L. Wagner, MSc; Maria J. Schymura, PhD

Poster Award Winners from the 2021 NAACCR Virtual Annual Conference
Pilot Study on Early Onset of Colorectal Cancer in Patients Under Age 50: How and Why are they Diagnosed? 138
Bobbi Jo Matt, MS, RHIT, CTR; Mary E. Charlton, PhD; Richard Hoffman, MD, MPH

Do Modifiable Risk Factors Impact Pancreatic Cancer Survival in a Population-based Study after Adjusting for Clinical Factors? 141
Mei-Chin Hsieh, PhD, CTR; Lu Zhang, PhD; Cruz Velasco-Gonzalez, PhD; Yong Yi, PhD; Lisa A. Pareti, BS, CTR; Edward J. Trapido, ScD, FACE; Vivien W. Chen, PhD; Xiao-Cheng Wu, MD, MPH, CTR

Late-Stage Cervical Cancer Diagnosis in Young Adults in California following the Affordable Care Act 144
Julianne J. P. Cooley; Frances B. Maguire; Renata Abrahão; Cyllene R. Morris; Arti Parikh-Patel; Theresa H. M. Keegan

Fall 2021 Continuing Education Quiz 146
Recinda Sherman, PhD, CTR

Call for Papers 147

Information for Authors 148

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.

Letters from the Editors

Dear Colleagues,

I am excited to announce another special edition of the Journal of Registry Management (JRM), created in partnership with the North American Association of Central Cancer Registries (NAACCR). We first partnered last year with NAACCR and have decided to continue that partnership with an annual NAACCR edition. Recinda Sherman, PhD, CTR, returns as guest editor. Recinda has worked in cancer surveillance for over 20 years, and her role at NAACCR is to support the use of central cancer registry data, including the NAACCR Cancer in North America (CiNA) data sets. The NAACCR edition also produced the winner of the 2020 Best Paper Award.

Regards,
Danette A. Clark, BS, RMA, AAS, CTR
Editor-in-Chief, *Journal of Registry Management*
JRMeditor@NCRA-USA.org

Dear Readers,

On behalf of NAACCR's Research and Data Use Steering Committee, I am pleased to present the second annual 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, and I encourage registries to plan to submit articles again next fall. I would like to congratulate again Dr. Maguire, UC Davis, and coauthors for winning the 2020 Journal of Registry Management Best Paper Award for their article (Treatment Patterns and Survival in Older Adults with Diffuse Large B-Cell Lymphoma: A Population-Based Study) in our inaugural NAACCR special edition.

This year's special issue includes 6 original articles. The papers presented underwent a peer review process overseen by NAACCR Research and Data Use Steering Committee Members. Accepted papers include a Canadian submission on adherence to diet and physical activities recommendations and reduced cancer risk (Wang et al). The remaining 5 papers are from the United States and include an article on the challenges of medical record abstraction in epidemiologic research (Ahmed et al); an article on selection bias introduced from relying on ePath to support rapid case ascertainment for patient contact studies (Kuliszewski et al); an article on survival predictors for pleural mesothelioma (Liu et al); an article investigating whether women with higher utilization of health care prior to diagnosis have better ovarian cancer outcomes (Kuliszewski et al); and an article on a demonstration project linking cancer registry data with administrative data on refugees, also the subject of the CE quiz (Morawski et al).

This issue also presents the winning posters from the NAACCR's 2021 Virtual Conference Poster Session. Poster authors participated in the Call for Abstracts for the NAACCR Summer Forum held in June 2021. After initial peer review, submissions accepted as a poster had the opportunity to include their posters in an online session. All judging was also conducted online, and, ultimately, 3 posters were selected for awards.

This issue includes the third-place winning poster, which documents changes in stage at diagnosis for cervical cancer in California following the passage of the Affordable Care Act (Cooley et al); the second-place winning poster, which evaluates which modifiable risk factors are important in pancreatic cancer outcomes (Hsieh et al); and the first-place winning poster that examines factors related to how colorectal cancer is diagnosed in patients under 50 years old (Matt et al).

It has been a privilege to collaborate with JRM on this second 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
Program Manager, Data Use & Research, NAACCR
Guest Editor, *Journal of Registry Management*

Original Article

Challenges of Medical Record Abstraction in a Long-Term Follow-up Study

Muhammad F. Ahmed, MBBS, MPH[a,b]; Andrea Galfo, MPH[a,b]; Wendy Huggins, BS[a,b]; Lisa E. Paddock, PhD, MPH[b,c]; Antoinette M. Stroup, PhD[b,c]; Jyoti Malhotra, MD, MPH[b]

Abstract: Medical records are a rich source of information and have tremendous value in epidemiological research. Nevertheless, the process of obtaining and abstracting medical records for a long-term follow-up study is complicated, time-consuming, and resource intensive. We identified the following major challenges during this process. First, widely varying infrastructure of electronic health record systems used by different organizations makes it difficult to ensure that all medical charts from all sources for a particular patient have been received. Second, extensive use of free text by health care providers requires a manual line-by-line search for relevant information, which may result in some missing data due to human error. Third, there are often discrepancies between patients' provided lists of health care providers and the registry data, which may affect the data-collection process. Fourth, providers have varied requirements for medical record release of their patients, which might entail multiple patient contacts. This, in turn, can frustrate patients and discourage them from participating in current or future research studies. Fifth, the use of inconsistent medical terminology by different providers complicates conversion of unstructured text into categorical data for analysis. We have the following recommendations for any future study with similar design to overcome the above challenges. First, the source of medical records best suited for the research objectives should be identified from the beginning. Second, the abstractors should be appropriately trained to accomplish research-specific tasks. Third, a quality data-tracking system for the abstracted elements should be employed to ensure data integrity. Fourth, the abstracted cases should be reviewed by one other abstractor. We also recommend a pilot study with a smaller number of patients to evaluate the required resources before any large-scale study.

Key words: *chart abstraction, electronic health records, long-term study, medical record abstraction*

Introduction

Medical records are rich in information and can be an invaluable resource in epidemiological research.[1-3] They are considered complementary to randomized controlled trials and health services research because of the availability of detailed clinical information on diagnosis, disease course, and treatment that may not be found elsewhere.[4-6] The US Food and Drug Administration has formally incorporated the electronic health record (EHR) as a source of real-world data to be used in research and has provided comprehensive guidelines for its use.[7] The enormously increased and remarkably improved use of computers in health care has furthered the use of medical records in research.[6,7]

The medical record abstraction (MRA), also known as *chart review*, is a process in which a human manually searches through an electronic or paper medical record to identify data required for a secondary purpose.[8] Notwithstanding the improvements, this process in a long-term study is complicated, time-consuming, and resource intensive.[6] The medical records are primarily collected for clinical purposes and are largely unstructured.[8,9] Subjective human inferences are often necessary when converting this raw data into categorical information for analysis. In this article, we will discuss the observations and challenges our team experienced during the collection and abstraction of medical records in a long-term follow-up study.

Method

We conducted MRA as a part of an observational study, "Identifying Racial Disparities in Follow-up Care in a Diverse Population of Lung Cancer Survivors (The Diversity Study)." The purpose of the study was to measure any racial differences among lung cancer survivors in receipt of the recommended posttreatment follow-up care, such as regular surveillance scans.

We identified 552 lung cancer survivors (189 males, 363 females) through the New Jersey State Cancer Registry (NJSCR) who met the study eligibility criteria. Each patient was mailed a research packet that included a cover letter explaining the purpose of the contact, a brochure of frequently asked questions about the study, a paper-based survey, a medical record release form to sign, and a form for providing a list of health care providers involved in the patient's lung cancer care.

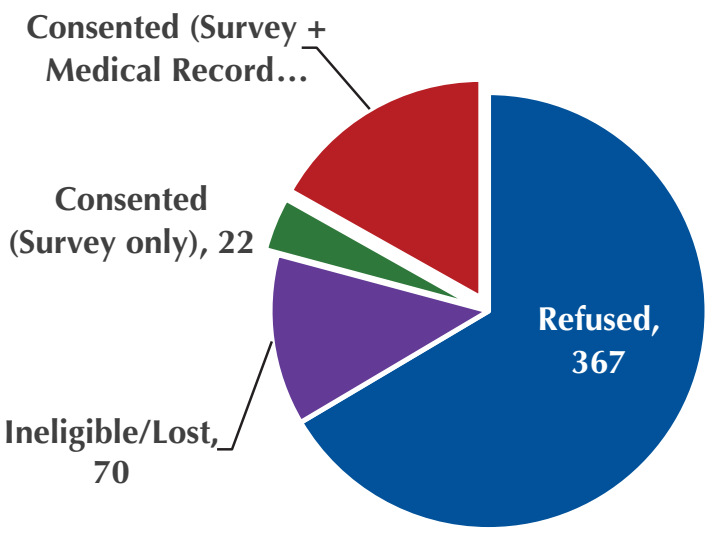
A total of 115 participants (20.8%) completed the paper-based survey. Of these, 93 (80.9%) returned a signed medical record release form (Figure 1). We then reached out to the physicians of the consented participants through mail to

[a] New Jersey Department of Health, Trenton, New Jersey. [b] Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey. [c] Rutgers School of Public Health, Piscataway, New Jersey.

Address correspondence to Muhammad F. Ahmed, MBBS, MPH. Email: Muhammad.Ahmed@doh.nj.gov.

Cancer Epidemiology Services, including the New Jersey State Cancer Registry, receives support from the National Program of Cancer Registries, Centers for Disease Control and Prevention under cooperative agreement NU58DP006279-02-00, the State of New Jersey, the National Cancer Institute, and the Rutgers Cancer Institute of New Jersey.

Figure 1. Outcome of Patient Contact



obtain the patients' medical records. A total of 261 records (an average of about 3 records per patient) were received, containing 5 years of posttreatment follow-up data for the 93 patients. These records were obtained from 150 facilities and 111 physician offices, following procedures compliant with the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Although the original charts were in electronic format, the charts received by the research staff were either mailed in a paper-based format or electronically faxed in a PDF format directly to the registry. In both situations, they required page-by-page or screen-by-screen manual review of data. We estimated that more than 25,000 pages were received and abstracted in total, for an average of about 300 pages of records per patient. A predesigned form was used to identify the data elements required for the study. These data were then entered in a Microsoft Access database that was created for this purpose. Additional data were collected from the NJSCR and patient-administered surveys (Table 1).

Quality-Control Measures

A team of 3 trained investigators abstracted data from the medical records. To ensure consistency, the investigators initially abstracted 3 cases together. Another combined session was held after 5 to 7 individual case abstractions to adopt the best practices from our respective experiences. Data from the first 5 cases were also reviewed by the principal investigator to ensure that we were capturing complete and accurate data using the medical record abstraction form. Thirty-five percent of the medical records (33 cases) were reabstracted by different staff for quality control.

Observations and Challenges

MRA for research studies may pose widely different challenges depending on the study objectives. For the diversity study, our team had to collect posttreatment follow-up data such as radiological tests, including chest radiographs, computed tomography (CT) scans, positron emission tomography (PET) scans, and magnetic resonance

Table 1. Sources of Information for the Diversity Study

Data type	*Source*
Basic demographic characteristics	NJSCR
Tumor characteristics	NJSCR
Health and social behaviors	Patient-administered survey
Comorbidities	Patient-administered survey
Treatment procedures and sequence	Medical records and NJSCR
Testing procedures, including radiographs and CT/PET scans	Medical records

CT, computed tomography; NJSCR, New Jersey State Cancer Registry; PET, positron emission tomography.

imaging (MRI). We analyzed the data as it was abstracted and tailored our model accordingly until the study objectives were achieved. This approach has also been evaluated and recommended by Polnaszek et al, who recommended a phase-based approach to find a "fit-for-use" framework for MRA.[10] The paragraphs below describe the major observations and challenges our team experienced during this process and the approach we took to succeed.

Ensuring a complete medical record is challenging with the complexity of EHR systems. The EHR landscape in health care is complex and subject to continuous and rapid changes.[11,12] Providers use a myriad of EHR systems with diverse configurations. This widely varied system of medical record repositories makes it difficult for researchers to determine if all required medical records from all years of follow-up for a particular patient have been received. To ensure complete records, our team requested and obtained an EHR from all possible sources for each patient, unless we determined that the already-received records had sufficient information for the study objectives. Although this required more resources, it provided a more accurate picture of posttreatment care.

Medical record release authorization: rigorous documentation requirements can frustrate patients. Providers have varied and rigorous requirements for the release of their patients' medical records, which often requires study staff to make multiple patient contacts throughout the recruitment process. This can frustrate patients and can discourage them from participating in current or future research studies. One possible solution to this issue is to provide a comprehensive medical records release form that includes language that is required by several providers. Additionally, the research staff should complete as many fields on the form as possible and leave blank only those requiring direct participant input before mailing it to them for a signature. This will not only save the participants from filling in the painstaking details required by many providers, but will also reduce the chance of errors.

Large/long medical records are time consuming and cumbersome to abstract. Large medical records can make data collection more difficult. We estimated that it took 1 abstractor approximately 1 hour to abstract 100 pages of free text. Therefore, large medical records (≥500 pages) posed a

substantial challenge to the abstractors. Due to the tedious and monotonous nature of the abstraction work, human factors like fatigue can affect the process. The legibility of some records is also compromised during photocopying and faxing, which may further complicate abstraction procedures. For the diversity study, we prioritized abstracting records related to testing procedures that were recorded over several years, because we found that they were more likely to represent the needed follow-up data rather than the extensive surgery-related notes.

Obtaining medical records specific to the study objectives needs careful consideration. Medical records contain different sets of information depending on their source, such as general hospitals, specialty centers, primary care providers, or subspecialty clinics. Determination of the sources that are best for the study objectives may require early abstraction and careful analysis of the first few records that are received. This should be reviewed with a provider before abstraction begins so that MRA goals are realistic. For instance, medical records from a hospital might be easier to obtain compared to a physician office.

Reporting bias: discrepancies in the list of physicians. We found that the patient-provided lists of physicians who treated their lung cancer were different from data in the NJSCR in 42% of cases. In the diversity study, we used patient-provided information with the NJSCR data in 89 out of 93 patients (95.7%). In the remaining 4 patients (4.3%), NJSCR data were used alone because the patients had not provided the lists. Although we expect minimal reporting bias in our study, this factor might affect the studies that use a single source for data collection.

Inconsistent medical terminology and extensive use of free text by providers. The use of inconsistent medical terminology by different providers complicates the conversion of unstructured text into categorical data for analysis. Furthermore, the providers extensively use free text, which requires a manual line-by-line search for relevant information. This may result in missing some data due to human error.

Obtaining a large number of medical records – 2-step strategy: Obtaining medical records from many sources requires a substantial amount of time, effort, and careful coordination. In our estimate, a single medical record from 1 source may take 8 to 10 hours of deliberations involving phone calls, faxing, and follow-ups, as well as receiving and scanning the records. We divided the process into 2 stages. In the beginning of the project, we conducted a mass mailing to all the health care providers that were identified by study participants. We were successful in receiving almost half of the records through this process. In the second stage, we reached out to the "nonresponding" health care providers through a more involved process, including individual phone calls and faxing. We received medical records of all 93 consented participants by the end of the study period.

Recommendations

We have the following recommendations for any future study involving MRA on long-term follow-up outcomes.

- *Source of research data:* Different sources of research data, such as medical records from various providers, registry data, and survey questionnaires, will provide different sets of information. The source that is best suited for the most complete and accurate outcome data should be determined early in the study. For instance, self-reported data is considered the reference standard for demographic information; however, registry data may be more accurate for clinical information.
- *Abstractor training:* To have consistency in the abstraction procedures, all staff should receive sufficient training from an experienced abstractor before starting the process. Other studies have also shown significant reduction in errors following a didactic training prior to the start of MRA.[13,14]
- *Tailoring the abstraction to research objectives:* Medical records often contain a tremendous amount of information. The researchers might be tempted to collect more information than required while designing the study or during the abstraction process. However, we recommend that the abstractors focus on the relevant data to save time and effort.
- *Data-collection tool:* By involving an experienced researcher, clinician, or medical records abstractor, a useful data-collection tool can be designed. This tool should be well tested before using it for the MRA.
- *Data abstraction audit:* A quality data tracking system for the abstracted elements might be required to ensure data integrity. Detailed reporting tools are helpful for tracking and organization.
- *Data quality-control measures:* We found that the cases abstracted by 2 abstractors independently had on the average 12.4 testing procedures per patient compared to 11.5 testing procedures per patient for the cases abstracted by 1 abstractor only. Therefore, we recommend that more than 1 abstractor separately review some or all charts depending on the available research resources. If possible, metrics like interrater reliability might also be used to measure the quality of abstracted data.[15]
- *Pilot study:* A pilot study with a smaller number of patients is strongly recommended to evaluate the required resources before any large-scale study is conducted.

Conclusion

Despite being a rich source of information, several factors can affect the data-collection process from medical records and thus bias research results. These include receiving incomplete medical records, inaccurate coding, and missing important information during abstraction. In our experience, the process can be improved by using multiple sources to identify the providers, adding a second abstractor for MRA, and analyzing the abstracted data early. It is also recommended to appropriately train the staff to obtain and abstract data, employ firm data auditing procedures, and allocate sufficient time and human resources to collect quality data to achieve the research objectives.

Acknowledgment

We would like to acknowledge the support of Rutgers Cancer Institute of New Jersey and the American Cancer Society for providing the funding for this study. We also thank all the hospitals and physician offices who provided the medical records for this study. A special thanks to all the hospital and central registry staff for their hard work in collection of cancer information.

References

1. Eder C, Fullerton J, Benroth R, Lindsay SP. Pragmatic strategies that enhance the reliability of data abstracted from medical records. *Appl Nurs Res*. 2005;18(1):50-54.
2. Gareen IF, Sicks JD, Jain AA, Moline D, Coffman-Kadish N. Identifying and collecting pertinent medical records for centralized abstraction in a multi-center randomized clinical trial: the model used by the American College of Radiology arm of the National Lung Screening Trial. *Contemp Clin Trials*. 2013;34(1):36-44.
3. Cook EA, Schneider KM, Robinson J, et al. Field methods in medical record abstraction: assessing the properties of comparative effectiveness estimates. *BMC Health Serv Res*. 2014;14:391.
4. Townsend JS, Jones MC, Jones MN, Waits AW, Konrad K, McCoy NM. A case study of early-onset colorectal cancer: using electronic health records to support public health surveillance on an emerging cancer control topic. *J Registry Manag*. 2021;48(1):4-11.
5. Ramagopalan SV, Simpson A, Sammon C. Can real-world data really replace randomised clinical trials? *BMC Med*. 2020;18(1):13.
6. Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. The evolving use of electronic health records (EHR) for research. *Semin Radiat Oncol*. 2019;29(4):354-361.
7. Real-world data (RWD) and real-world evidence (RWE) are playing an increasing role in health care decisions. US Food and Drug Administration website. Accessed August 3, 2021. **https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence**
8. Zozus MN, Pieper C, Johnson CM, et al. Factors affecting accuracy of data abstracted from medical records. *PLoS One*. 2015;10(10):e0138649.
9. Allison JJ, Wall TC, Spettell CM, et al. The art and science of chart review. *Jt Comm J Qual Improv*. 2000;26(3):115-136.
10. Polnaszek B, Gilmore-Bykovskyi A, Hovanes M, et al. Overcoming the challenges of unstructured data in multisite, electronic medical record-based abstraction. *Med Care*. 2016;54(10):e65-e72.
11. Jiancaro T, Jamieson GA, Mihailidis A. Twenty years of cognitive work analysis in health care: a scoping review. *J Cogn Eng Decision Making*. 2014;8(1):3-22.
12. St-Maurice JD, Burns CM. Modeling patient treatment with medical records: an abstraction hierarchy to understand user competencies and needs. *JMIR Hum Factors*. 2017;4(3):e16.
13. Zozus MN, Young LW, Simon AE, et al. Training as an intervention to decrease medical record abstraction errors multicenter studies. *Stud Health Technol Inform*. 2019;257:526-539.
14. Mi MY, Collins JE, Lerner V, et al. Reliability of medical record abstraction by non-physicians for orthopedic research. *BMC Musculoskelet Disord*. 2013;14:181.
15. Roth CP, Coulter ID, Kraus LS, et al. Researching the appropriateness of care in the complementary and integrative health professions part 5: using patient records: selection, protection, and abstraction. *J Manipulative Physiol Ther*. 2019;42(5):327-334.

Original Article

Association between Cancer Risk and Behaviors Adherent to Cancer Prevention Recommendations in Ontario, Canada

Ying Wang, MSc[a]; Mohammad Haque, MPH[a]; Stephanie Young, MPH[a]; Michelle Cotterchio, PhD[a, b]; Rebecca Truscott, MHSc, RD[a]

Abstract: Purpose: In 2007, the World Cancer Research Fund (WCRF) and American Institute for Cancer Research (AICR) published several diet and physical activity recommendations to reduce cancer risk. Our objective was to examine the association between self-reported behaviors consistent with the WCRF/AICR recommendations and the risk of developing any cancer and colorectal cancer in Ontario. Method: 111,139 Ontarians who completed the Canadian Community Health Survey (2000–2008) were linked to the Ontario Cancer Registry to determine whether they were diagnosed with cancer. Their responses were used to assess behaviors consistent with 4 WCRF/AICR recommendations (body fatness, physical activity, vegetable and fruit consumption, and alcoholic drinks). Multivariate Cox proportional hazard regression models were used to assess the association between adherence to the 4 WCRF/AICR recommendations and subsequent cancer risk. Results: Among the 111,139 participants, 8,942 (8%) were diagnosed with cancer with a mean follow-up of 9.6 years. Compared to not meeting any of the selected WCRF/AICR recommendations (composite score, 0), participants who were most adherent to the selected WCRF/AICR recommendations (composite score, 4) were 31% less likely to develop any cancer (HR, 0.69; 95% CI, 0.51–0.92) and were 61% less likely to develop colorectal cancer (HR, 0.39; 95% CI, 0.20–0.77) after adjusting for some potential confounding factors. When stratified by sex, the associations remained statistically significant for men, but not for women. In addition, increasing vegetable and fruit consumption, having lower body fatness, and decreasing alcohol consumption were each associated with reduced risk of both any cancer and colorectal cancer. Conclusion: Healthy behaviors consistent with select WCRF/AICR recommendations were associated with a decreased risk of developing any cancer and colorectal cancer among this Ontario cohort.

Key words: *cancer incidence, cancer prevention recommendations, healthy behavior, linkage*

Introduction

In 2015, more than 81,000 new cases of cancer were diagnosed in Ontario, with the most common sites being breast, colorectal, and lung. The same year, more than 29,000 people died of cancer in Ontario, making it the most common cause of death in the province.[1,2] It is estimated that, in 2015, between 33% and 37% of new cancer cases among Canadian adults aged ≥30 years were attributable to preventable risk factors.[3] Of these risk factors, tobacco smoking, physical inactivity, and excess body weight were responsible for the highest proportions of preventable cases.[3-6]

Certain individual behaviors and conditions (eg, alcohol intake, obesity) are known to increase the risk of some cancer types, while others may reduce cancer risk (eg, physical activity). In 2007, the World Cancer Research Fund (WCRF) and the American Institute for Cancer Research (AICR) published a report that summarized the epidemiological evidence relating food, nutrition, physical activity, and body composition to the risk of cancer, as well as cancer prevention recommendations related to body fatness, physical activity, and diet.[7] Since these recommendations were published, studies in other countries and provinces have reported that increased adherence to the WCRF/AICR recommendations was associated with decreased overall cancer risk[8-11] as well as a reduction in the rate of colorectal cancer.[10-15]

This study examined the association between self-reported behavior consistent with the WCRF/AICR cancer prevention recommendations for body fatness, physical activity, vegetable and fruit consumption, and alcoholic drinks and the risk of developing cancer in a population-based cohort from Ontario, Canada. Colorectal cancer was also examined, as this is one of the most commonly diagnosed cancers in Ontario, and it is associated with at least 2 of the WCRF/AICR behaviors examined in this study (obesity, alcohol, and dietary intake).[7]

Methods

Population-Based Record Linkage Cohort

The main study cohort was derived from the Ontario sample of the Canadian Community Health Survey (CCHS)—specifically, cycles 1.1 (2000–2001), 2.1 (2003), 3.1 (2005), and 2007–2008 obtained from the Ontario Ministry of Health and Long-Term Care. The CCHS is a cross-sectional population-based survey conducted by Statistics

[a]Ontario Health (Cancer Care Ontario), Toronto, Ontario, Canada. [b]Dalla Lana School of Public Health, University of Toronto, Ontario, Canada.

Address correspondence to Ying WANG, Population Health and Prevention, Ontario Health/Cancer Care Ontario, 505 University Avenue, Toronto, Ontario, M5G 1X3, Canada. Email: ying.wang@ontariohealth.ca.

Canada that collects information related to the health status, health care use, and determinants of health for the Canadian population aged ≥12 years.[16] Specifically, 159,474 (93.4%) of Ontario CCHS respondents consented to have their information shared with the province and linked with administrative health databases, which allowed this study's cohort data to be linked to other databases for cancer incidence and mortality information by their unique health insurance number.

To obtain information on cancer incidence, the study cohort was linked to the Ontario Cancer Registry using health insurance number. This registry is a computerized database containing information on all Ontario residents who have been diagnosed with invasive neoplasia (except for basal cell and squamous cell skin cancers) since 1964.[17] Major data sources of the Ontario Cancer Registry include the following: cancer-related hospital discharge and day surgery reports from the Canadian Institute for Health Information; cancer-related pathology reports, received mostly electronically from hospital and community laboratories; consultation and treatment records of patients referred to 1 of 14 regional cancer centers; and death certificates with cancer identified as the underlying cause of death received from the Ontario Registrar General.[17]

To obtain mortality information, the study cohort was linked to the Registered Persons Database using health insurance number. The Registered Persons Database contains demographic information (sex, age, date of birth, vital status, date of death) for all Ontario residents who are eligible for health care coverage through the Ontario Health Insurance Plan.

After the linkage of the CCHS, the Ontario Cancer Registry, and Registered Persons Database files, 134,567 respondents with valid health insurance numbers were identified from year 2000 to 2008. If a respondent was included in multiple cycles of the CCHS, only information collected from their earliest survey response was used to maximize the follow-up period. Also, only 0.8% of participants in the 4 cycles of CCHS survey have repeated measurements, so time-dependent covariates could not be incorporated into the analysis due to the data limitation. Thus, 1,040 repeated measurements were removed. Several other exclusion criteria were applied. Respondents were excluded if they were younger than 18 years at the time of survey administration (n = 11,770), had a past diagnosis of cancer (n = 6,708), were pregnant at time of interview (n = 1,129), became ineligible for Ontario Health Insurance Plan during the follow-up period (n = 6), had a body mass index (BMI) less than 18.5 kg/m^2 (n = 2,770), or where data quality was a concern (n = 5); for example, if the death date was earlier than the interview date. A final sample of 111,139 respondents was identified for the analysis of all cancers and a sample of 103,285 participants was identified for colorectal cancer analysis (7,854 patients diagnosed with cancer types other than colorectal cancer were removed).

Descriptive statistics were generated to examine demographic characteristics of the cross-sectional survey samples. The percentage within groups for categorical variables, as well as group means and standard deviations for continuous variables, are reported in Table 1.

Cancer Prevention Recommendations Adherence Score Operationalization

The 8 cancer prevention recommendations published in the 2007 WCRF/AICR report were related to body fatness, physical activity, foods and drinks that promote weight gain, plant foods, food preservation, processing and preparation, animal foods, alcoholic drinks, and dietary supplements.[7] In the CCHS, data were available to examine 4 of these recommendations:

1. *Body fatness:* Be as lean as possible without becoming underweight
2. *Physical activity:* Be physically active as part of your everyday life
3. *Plant foods:* Eat mostly foods of plant origin
4. *Alcoholic drinks:* Limit alcoholic drinks

The WCRF/AICR cancer prevention recommendations were updated in 2018. However, the authors deemed measuring individual adherence to the 2007 WCRF/AICR recommendations was more appropriate given this study's cohort was followed up to 2014, prior to the updated 2018 recommendations.

To quantify the level of adherence to the cancer prevention recommendations, we created an adherence score based on respondents' behaviors. For each of the 4 recommendations listed above, survey respondents were assigned a value of 1 when the recommendation was fully met, 0.5 when it was partially met, and 0 when it was unmet (Table 2). An aggregate WCRF/AICR composite adherence score was created by summing the 4 individual scores. The composite adherence score ranged from 0 to 4, with a higher score representing behaviors more consistent with the WCRF/AICR cancer prevention recommendations. Details about how the WCRF/AICR recommendations were operationalized using variables from the CCHS are provided in Table 3, and stated in brief below:

1. *Body fatness:* The subrecommendation, "maintain body weight within the normal range," was evaluated based on a respondent's self-reported height and weight to calculate their BMI (weight [kilograms]/height [meters]^2). Adherence to this recommendation was defined as those with a BMI between 18.5 and 24.99 kg/m^2, while those with a BMI > 24.99 kg/m^2 did not adhere, and those with a BMI < 18.5 kg/m^2 were excluded from the analyses.
2. *Physical activity:* The subrecommendation, "be moderately physically active" every day was measured by daily energy expenditure (EE), where adherence was defined as EE ≥ 3.0 kcal/kg/d, partial adherence was 1.5 ≤ EE < 3.0 kcal/kg/d, and nonadherence was EE <1.5 kcal/kg/d.
3. *Plant foods:* The subrecommendation, "eat ≥ 5 portions/servings (≥400 g)" every day was measured by daily vegetable and fruit consumption, where adherence was defined as eating nonstarchy vegetables and fruit ≥ 5 times a day, partial adherence was 3–4 times a day, and nonadherence was ≤ 2 times a day.
4. *Alcoholic drinks:* The recommendation to "limit consumption to ≤ 2 drinks per day for men and 1 drink per day for women" was measured by average daily

Table 1. Baseline Characteristics of Study Participants for Cases and Controls by Cancer Cohort						
Baseline characteristic/category	*Overall cohort (N = 111,139)*			*Colorectal cancer cohort (N = 103,285)*		
	Cases (n = 8,942)	**Controls (n = 102,197)**	***P* value [a]**	**Cases (n = 1,088)**	**Controls (n = 102,197)**	***P* value [a]**
Continuous variables, mean (SD)						
Age, y	63 (13)	48 (18)	<.0001	67 (13)	48 (18)	<.0001
Body mass index, kg/m^2	27 (5)	26 (5)	.1125	27 (5)	26 (5)	.0334
Categorical variables, n (%)						
Sex			<.0001			.0085
Male	4,426 (50)	47,422 (46)		548 (50)	47,422 (46)	
Female	4,514 (50)	54,757 (54)		539 (50)	54,757 (54)	
Missing	2 (0)	18 (0)		1 (0)	18 (0)	
Education			<.0001			<.0001
Less than secondary school	81 (1)	19,795 (19)		399 (37)	19,795 (19)	
Secondary school graduate	2,796 (31)	20,412 (20)		192 (18)	20,412 (20)	
Some postsecondary	1,578 (18)	7,966 (8)		60 (5)	7,966 (8)	
Postsecondary graduate	536 (6)	53,293 (52)		430 (40)	53,293 (52)	
Missing	3,951 (44)	731 (1)		7 (0)	731 (1)	
Ethnicity			<.0001			<.0001
White	7,626 (85)	81,809 (80)		929 (85)	81,809 (80)	
Nonwhite	420 (5)	9,537 (9)		45 (4)	9,537 (9)	
Missing	896 (10)	10,851 (11)		114 (11)	10,851 (11)	
Marital status			<.0001			<.0001
Married or common-law	5,476 (61)	58,814 (58)		636 (58)	58,814 (58)	
Widowed, separated, or divorced	2,752 (31)	19,781 (19)		379 (35)	19,781 (19)	
Single or never married	712 (8)	23,563 (23)		72 (7)	23,563 (23)	
Missing	2 (0)	39 (0)		1 (0)	39 (0)	
Household income (Canadian dollars)			<.0001			<.0001
No income to $29,999	2,713 (30)	21,340 (21)		373 (34)	21,340 (21)	
$30,000 to $49, 999	2,040 (23)	19,286 (19)		238 (22)	19,286 (19)	
$50,000 to $79, 999	1,892 (21)	24,623 (24)		219 (20)	24,623 (24)	
$80,000 to $99, 999	1,075 (12)	17,812 (17)		106 (10)	17,812 (17)	
$100,000 or more	446 (5)	10,754 (11)		38 (3)	10,754 (11)	
Missing	776 (9)	8,382 (8)		114 (11)	8,382 (8)	
Immigrant status			<.0001			.237
Immigrant	1,967 (22)	20,415 (20)		238 (22)	20,415 (20)	
Canadian-born	6,962 (78)	81,663 (80)		848 (78)	81,663 (80)	
Missing	13 (0)	119 (0)		2 (0)	119 (0)	
Residence (geography)			<.0001			<.0001
Urban	6,910 (77)	80,840 (79)		804 (74)	80,840 (79)	
Rural	2,032 (23)	21,357 (21)		284 (26)	21,357 (21)	
Smoking status			<.0001			<.0001
Current smoker	2,167 (24)	25,614 (25)		186 (17)	25,614 (25)	
Former smoker	4,475 (50)	42,181 (41)		581 (53)	42,181 (41)	
Never smoker	2,291 (26)	34,296 (34)		1 (0)	106 (0)	
Missing	9 (0)	106 (0)		1 (0)	106 (0)	

[a] *P* values for the comparison between cases/noncases using χ^2 tests or Fisher tests as appropriate. All statistical tests were 2-sided.

alcohol consumption in the past week, where adherence was defined as consuming ≤ 2 drinks per day for men and ≤1 drink/day for women, partial adherence was > 2 to 3 drinks/day for men and >1 to 2 drinks for women, and nonadherence was > 3 drinks per day for men and >2 drinks per day for women.

Incidence of Cancer

The outcome of interest for this study was the development of incident invasive cancer. The cancer types are defined by the International Classification of Diseases for Oncology 3rd edition (ICD-O-3) topography codes from the US Surveillance, Epidemiology, and End Results (SEER) Program's Site Recode definition.[18] The primary outcome of invasive cancer was defined as ICD-O-3 topography codes C00.0–C80.9 and behavior code of 3. The secondary outcome of interest was the development of incident colorectal cancer (ICD-O-3 topography codes C18–C19 and C26.0). Study follow-up for every individual spanned from the date of their survey administration (from 2000 to 2008) through to December 31, 2014. Survey respondents were followed until the earliest of the following events: diagnosis of invasive cancer, death, or the end of the study period (censored).

Statistical Analysis

The association between behaviors consistent with the WCRF/AICR cancer prevention recommendations (adherence score) and the risk of developing any cancer and specifically colorectal cancer were assessed. We tested the association with incident cancer risk for each individual health behavior (individual score) and overall behavior (composite score) for both sexes combined and males and females separately. We applied multivariate Cox proportional hazards regression models to assess the association between adherence to the 4 WCRF/AICR recommendations and subsequent cancer risk. The Cox proportional hazards model is a widely used semiparametric model in the analysis of survival data to explain the effect of explanatory variables on hazard ratios (HRs). *Time to event* was measured from the date of CCHS survey completion to the date of cancer diagnosis. Individuals who did not develop cancer during the follow-up period were censored at the date of death or at the end of the study period (December 31, 2014) as appropriate. The PHREG procedure in the statistical analytic software SAS 9.4 was used to calculate the HR with corresponding 95% CI. SAS is a command-driven software package used for statistical analysis and data visualization.[19]

Table 2. Proportions of Participants Meeting Selected WCRF/AICR Recommendations by Sex and Cancer Cohort

WCRF/AICR recommendations		*Overall cohort (N = 111,139)*			*Colorectal cancer cohort (N = 103,285)*		
Operationalization	Score[a]	Male (%)	Female (%)	Both sexes combined (%)	Male (%)	Female (%)	Both sexes combined (%)
Body fatness (BMI; kg/m^2)							
Normal (18.5–24.99)	1	39	50	45	39	50	45
Overweight and obese (>24.99)	0	60	47	53	60	46	53
Physical activity							
Active (EE ≥ 3.0 kcal/kg/d)	1	28	22	25	28	22	25
Moderately active (EE 1.5 – < 3.0 kcal/kg/d)	0.5	24	25	25	24	25	25
Inactive (EE < 1.5 kcal/kg/d)	0	45	52	49	45	51	48
Plant foods							
Adequate consumption of fruit and vegetable intake (≥5/d)	1	21	36	29	21	36	29
Partially adequate consumption of fruit and vegetable intake (3–4/d)	0.5	56	50	53	56	50	53
Inadequate consumption of fruit and vegetable intake (≤2/d)	0	20	12	15	20	12	15
Alcoholic drinks							
Nondrinkers (≤2/d for men and ≤1/d for women)	1	87	91	89	87	91	89
Moderate drinkers (>2 to 3/d for men and >1–2/d for women)	0.5	6	6	6	6	6	6
Heavy drinkers (>3/d for men and >2/d for women)	0	5	2	4	5	2	4

AICR, American Institute for Cancer Research; BMI, body mass index (kg/m^2); EE, energy expenditure; WCRF, World Cancer Research Fund.

[a]The WCRF/AICR recommendation score is the score assigned to each individual based on their health behavior adherence to 1 of the 4 WCRF/AICR cancer prevention recommendations examined in this study: body fatness, physical activity, vegetable and fruit consumption (used as a proxy for plant foods), and alcoholic drinks. Each individual score was quantified as 0 (no adherence), 0.5 (partial adherence) or 1 (full adherence) based on survey responses.

Table 3. Operationalization of WCRF/AICR Recommendations			
WCRF/AICR[7] recommendations	***Subrecommendations***	***Operationalization***	***Score[a]***
(1) Body fatness. Be as lean as possible without becoming underweight.	(1a) Ensure that body weight throughout childhood and adolescent growth projects toward the lower end of the normal BMI at age 21 years.	Insufficient data	–
	(1b) Maintain body weight within the normal range from age 21 years.	BMI 18.5–24.99 kg/m^2	1
		BMI < 18.5 or > 24.99 kg/m^2	
	(1c) Avoid weight gain and increases in waist circumference throughout adulthood.	Insufficient data	0
(2) Physical activity. Be physically active as part of your everyday life.	(2a) Be moderately physically active, equivalent to brisk walking, for ≥30 min every day.	Active (EE ≥ 3.0 kcal/kg/d)	1
		Moderately active (EE 1.5 – < 3.0 kcal/kg/d)	0.5
		Inactive (EE < 1.5 kcal/kg/d)	0
	(2b) As fitness improves, aim for ≥60 min of moderate activity or ≥30 min of vigorous physical activity every day.	Insufficient data	–
	(2c) Limit sedentary habits such as watching television.	Insufficient data	–
(3) Foods and drinks that promote weight gain. Limit consumption of energy-dense foods; avoid sugary drinks.	(3a) Consume energy-dense foods sparingly.	Insufficient data	–
	(3b) Avoid sugary drinks.	Insufficient data	–
	(3c) Consume fast foods sparingly, if at all.	Insufficient data	–
(4) Plant foods. Eat mostly foods of plant origin	(4a) Eat ≥5 portions/servings (≥400 g) of a variety of nonstarchy vegetables and of fruit every day.	Fruit and vegetable intake: ≥5/d	1
		Fruit and vegetable intake: 3–4/d	
		Fruit and vegetable intake: 3–4/d	0.5
		Fruit and vegetable intake: ≤2/d	0
	(4b) Eat relatively unprocessed cereals (grains) and/or pulses (legumes) with every meal.	Insufficient data	–
	(4c) Limit refined starchy foods.	Insufficient data	–
	(4d) People who consume starchy roots or tubers as staples should also ensure sufficient intake of nonstarchy vegetables, fruit, and pulses (legumes).	Insufficient data	–
(5) Animal foods. Limit intake of red meat and avoid processed meat.	(5a) People who eat red meat should consume <500 g/wk and very few, if any, processed meats.	Insufficient data	–
(6) Alcoholic drinks. Limit alcoholic drinks.	(6a) If alcoholic drinks are consumed, limit consumption to ≤2/d for men and 1/d for women.	≤2/d (men)	1
		≤1/d (women)	1
		>2–3/d (men)	0.5
		>1–2/d (women)	0.5
		>3/d (men)	0
		>2/d (women)	0
(7) Preservation, processing, preparation. Limit consumption of salt. Avoid moldy cereals (grains) or pulses (legumes).	(7a) Avoid salt-preserved, salted, or salty foods; preserve foods without using salt.	Insufficient data	–
	(7b) Limit consumption of processed foods with added salt to ensure an intake of <6 g (2.4 g sodium)/d.	Insufficient data	–
	(7c) Do not eat moldy cereals (grains) or pulses (legumes).	Insufficient data	–
(8) Dietary supplements. Aim to meet nutritional needs through diet alone.	(8a) Dietary supplements are not recommended for cancer prevention.	Insufficient data	–

AICR, American Institute for Cancer Research; BMI, body mass index (kg/m^2); EE, energy expenditure; WCRF, World Cancer Research Fund.
[a] Participants were scored 1 if they adhered to the recommendation, 0.5 if they partially adhered, and 0 if they did not adhere to the WCRF/AICR recommendations.

Table 4. Proportions of CCHS Participants Meeting Selected WCRF/AICR Recommendations for Cases and Controls by Cancer Cohort

WCRF/AICR recommendations operationalization		*Overall cohort (N = 111,139)*		*P value*[a]	*Colorectal cancer cohort (N = 103,285)*		*P value*[a]
Category	**Score**[b]	**Cases**	**Controls**		**Cases**	**Controls**	
Body fatness[c]				<.0001			<.0001
Overweight and obese	0.0	5,270 (59)	53,851 (53)		688 (63)	53,851 (53)	
Normal	1.0	3,472 (39)	46,079 (45)		378 (35)	46,079 (45)	
Physical activity[d]				<.0001			<.0001
Inactive	0.0	4,731 (53)	49,177 (48)		600 (55)	49,177 (48)	
Moderate active	0.5	2,171 (24)	25,431 (25)		251 (23)	25,431 (25)	
Active	1.0	1,808 (20)	25,381 (25)		213 (20)	25,381 (25)	
Plant foods[e]				.0002			.1059
Inadequate consumption	0.0	1,242 (14)	15,662 (15)		144 (13)	15,662 (15)	
Partially adequate consumption	0.5	4,868 (55)	53,817 (53)		601 (55)	53,817 (53)	
Adequate consumption	1.0	2,526 (28)	29,632 (29)		311 (29)	29,632 (29)	
Alcoholic drinks[f]				.0286			.0703
Heavy drinkers	0.0	291 (3)	3,740 (3)		40 (4)	3,740 (4)	
Moderate drinkers	0.5	492 (6)	6,114 (6)		47 (4)	6,114 (6)	
Non-drinkers	1.0	7,991 (89)	90,775 (89)		984 (90)	90,775 (89)	
Composite adherence score[g]				<.0001			<.0001
0.0 (no adherence)	0.0	56 (1)	475 (1)		11 (1)	475 (1)	
0.5 (partial adherence)	0.5	118 (1)	1,250 (1)		15 (1)	1,250 (1)	
1.0	1.0	819 (9)	8,575 (8)		101 (9)	8,575 (8)	
1.5	1.5	1,930 (22)	18,669 (18)		265 (24)	18,669 (18)	
2.0	2.0	1,889 (21)	20,461 (20)		210 (19)	20,461 (20)	
2.5	2.5	1,971 (22)	22,672 (22)		238 (22)	22,672 (22)	
3.0	3.0	1,230 (14)	15,963 (16)		144 (13)	15,963 (16)	
3.5	3.5	636 (7)	9,213 (9)		67 (6)	9,213 (9)	
4.0 (full adherence)	4.0	292 (3)	4,902 (5)		37 (3)	4,902 (5)	

Values are no. (%) unless otherwise indicated.

AICR, American Institute for Cancer Research; BMI, body mass index (kg/m^2); EE, energy expenditure; WCRF, World Cancer Research Fund.

[a] P values for the comparison between cases/noncases using χ^2 tests. All statistical tests were 2-sided.

[b] The WCRF/AICR Recommendation Score was created by assigning a value of 1 to each recommendation that was fully met, 0.5 to each recommendation that was partially met, and 0 to each recommendation that was unmet, then summing the scores for each recommendation for the composite adherence score.

[c] Body fatness is defined based on a person's BMI in kg/m^2: Overweight and obese (BMI ≥ 25), score 0; normal (18.5 ≤ BMI < 25), score 1.

[d] Physical activity is classed as inactive (EE < 1.5 kcal/kg/d), score 0; moderately active (1.5 kcal/kg/d ≤ EE < 3.0), score 0.5; active (EE ≥ 3.0 kcal/kg/d), score 1.

[e] Plant food is defined based on fruit and vegetable intake per day: inadequate (≤2/d), score 0; partially (3–4/d), score 0.5; adequate (≥5/d), score 1.

[f] Alcoholic drinking is defined based on daily consumption: heavy drinkers (>27.2 g/d), score 0; moderate drinkers (13.6–27.2 g/d), score 0.5; and nondrinkers (≤13.6 g/d), score 1.

[g] Composite adherence score is defined based on the 4 WCRF/AICR cancer prevention recommendations examined in this study, a total value ranging from 0 to 4 points: participants who were most adherent to the recommendations (composite score: 4) and individuals not meeting WCRF/AICR recommendations (composite score: 0).

Table 5. Associations Between Categories of the WCRF/AICR Adherence Score and Cancer Risk

WCRF/AICR recommendations operationalization	*Individual score*	*Both sexes*		*Male*		*Female*	
		HR[a]	**95% CI**	**HR[a]**	**95% CI**	**HR[a]**	**95% CI**
Any cancer[b]							
Body fatness							
Overweight and obese	0	1.00	ref	1.00	ref	1.00	ref
Normal	1	0.90	0.86–0.94*	0.87	0.82–0.93*	0.89	0.84–0.95*
P trend[c]		<.0001*		<.0001*		.0002*	
Physical activity							
Inactive	0	1.00	ref	1.00	ref	1.00	ref
Moderate active	0.5	0.99	0.94–1.04	0.93	0.86–1	1.01	0.94–1.09
Active	1	0.95	0.9–1	0.95	0.88–1.02	0.90	0.83–0.98*
P trend[c]		.1577		.1062		.0266*	
Plant foods							
Inadequate consumption	0	1.00	ref	1.00	ref	1.00	ref
Partially consumption	0.5	0.93	0.87–0.99*	0.89	0.82–0.97*	0.97	0.88–1.07
Adequate consumption	1	0.90	0.84–0.97*	0.84	0.76–0.92*	0.95	0.85–1.06
P trend[c]		.0157*		.0015*		.5875	
Alcoholic drinks							
Heavy drinkers	0	1.00	ref	1.00	ref	1.00	ref
Moderate drinkers	0.5	0.93	0.80–1.08	0.92	0.77–1.11	0.92	0.71–1.18
Nondrinkers	1	0.88	0.78–0.99*	0.82	0.71–0.94*	0.92	0.74–1.14
P trend[c]		.0594		.0060*		.7461	
Adherence composite score							
0.0 (no adherence)	0.0	1.00	ref	1.00	ref	1.00	ref
0.5 (partial adherence)	0.5	0.82	0.60–1.13	0.84	0.58–1.20	0.75	0.37–1.54
1.0	1.0	0.83	0.63–1.10	0.85	0.62–1.15	0.76	0.40–1.42
1.5	1.5	0.82	0.62–1.07	0.79	0.58–1.06	0.80	0.43–1.49
2.0	2.0	0.79	0.60–1.03	0.8	0.59–1.07	0.73	0.39–1.36
2.5	2.5	0.78	0.60–1.02	0.75	0.55–1.01	0.75	0.40–1.39
3.0	3.0	0.73	0.56–0.96*	0.72	0.53–0.98*	0.68	0.36–1.27
3.5	3.5	0.71	0.54–0.93*	0.68	0.50–0.93*	0.65	0.35–1.23
4.0 (full adherence)	4.0	0.69	0.51–0.92*	0.6	0.43–0.86*	0.65	0.34–1.23
P trend[c]		.0009*		.0036*		.0109*	

Table 5, cont. Associations Between Categories of the WCRF/AICR Adherence Score and Cancer Risk

WCRF/AICR recommendations operationalization	*Individual score*	*Both sexes*		*Male*		*Female*	
		HR[a]	95% CI	HR[a]	95% CI	HR[a]	95% CI
Colorectal Cancer[d]							
Body fatness							
Overweight and obese	0	1.00	ref	1.00	ref	1.00	ref
Normal	1	0.75	0.66–0.86*	0.73	0.61–0.88*	0.76	0.64–0.91*
P trend[c]		<.0001*		.001*		.0027*	
Physical activity							
Inactive	0	1.00	ref	1.00	ref	1.00	ref
Moderate active	0.5	0.91	0.78–1.05	0.79	0.64–0.98*	1.03	0.83–1.26
Active	1	0.90	0.77–1.05	0.76	0.62–0.95*	1.08	0.85–1.36
P trend[c]		.2765		.0169*		.8238	
Plant foods intake							
Inadequate consumption	0	1.00	ref	1.00	ref	1.00	ref
Partially consumption	0.5	0.87	0.72–1.05	0.73	0.58–0.92*	1.22	0.87–1.72
Adequate consumption	1	0.82	0.66–1	0.69	0.53–0.91*	1.11	0.78–1.58
P trend[c]		.1511	.0130*	.3672			
Alcoholic drinks							
Heavy drinkers	0	1.00	ref	1.00	ref	1.00	ref
Moderate drinkers	0.5	0.60	0.39–0.91*	0.72	0.43–1.19	0.50	0.22–1.12
Nondrinkers	1	0.66	0.48–0.9*	0.61	0.42–0.88*	0.77	0.4–1.49
P trend[c]		.0271*		.0228*		.1528	
Adherence composite score							
0.0 (no adherence)	0.0	1.00	ref	1.00	ref	1.00	ref
0.5 (partial adherence)	0.5	0.52	0.24–1.13	0.55	0.24–1.26	0.48	0.04–5.24
1.0	1.0	0.49	0.26–0.91*	0.52	0.27–1.01	0.61	0.08–4.50
1.5	1.5	0.51	0.28–0.94*	0.47	0.25–0.9*	0.85	0.12–6.10
2.0	2.0	0.40	0.22–0.74*	0.4	0.21–0.76*	0.63	0.09–4.51
2.5	2.5	0.43	0.24–0.79*	0.39	0.20–0.74*	0.73	0.10–5.22
3.0	3.0	0.39	0.21–0.72*	0.32	0.16–0.63*	0.69	0.10–4.98
3.5	3.5	0.34	0.18–0.64*	0.31	0.15–0.63*	0.56	0.08–4.08
4.0 (full adherence)	4.0	0.39	0.20–0.77*	0.27	0.12–0.63*	0.75	0.10–5.55
P trend[c]		.0030 *		.0027 *		.3214	

AICR, American Institute for Cancer Research; WCRF, World Cancer Research Fund.

* Statistically significant.

[a] Hazard ratio is adjusted with potential confounding factors age, sex, and smoking status.

[b] N = 111,139; male, n = 51,848; female, n = 59,271; case n = 8,942; male, n = 4,426, female, n = 4,514.

[c] Trend tested by modelling categories of adherence to the WCRF/AICR individual score as an ordinal variable ranged from 0 to 1; modelling categories of adherence to the WCRF/AICR composite score as an ordinal variable from 0 to 4 (sum of the individual adherence scores).

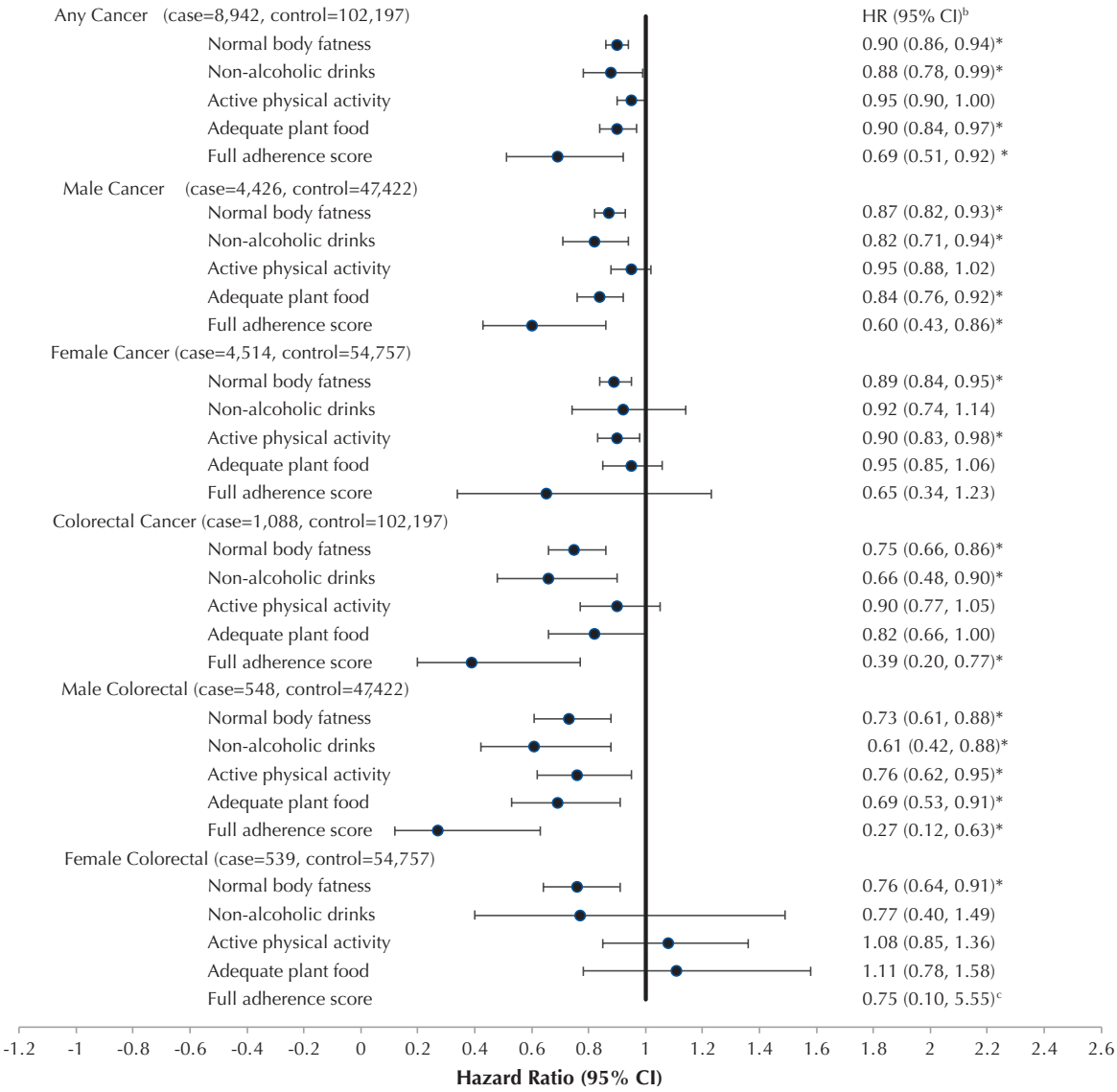
[d] N = 103,285; male, n = 47,970; female, n = 55,296; case n = 1,088, male, n = 548; female, n = 539.

Age, sex, smoking type, education, ethnicity, marital status, immigrant status, and geography residence were tested as potential confounding factors by computing the measure of association both before and after adjusting for a potential confounding factor sequentially. Advancing age, male sex, and tobacco use are the most common confounders for cancer overall and for many individual cancer types.[20-23] The assumptions of the Cox proportional hazards model were also assessed for some key independent variables and the Schoenfeld residuals were independent of time. Finally, age at baseline (measured in years as a continuous variable), sex (male and female), and smoking status (current, former, or never smokers) were retained as the covariates in the final multivariate Cox proportional hazard models.

Results

The baseline sociodemographic and health characteristics of participants in cases and controls for the overall cohort and colorectal cancer cohort are described in Table 1. Among the 111,139 participants in this study, 53% were female and the mean age at enrollment was 50 years (SD = 19); 47% were males and they had a slightly younger mean age of 48 years (SD = 18). From 2000 to 2014, a total of 8,942 overall incident cancers were identified (8% of participants) with a mean follow-up of 9.6 years, and 1,088 (1%) colorectal cancer were diagnosed with a mean follow-up of 9.9 years. The mean diagnosed age was 68 years (SD = 13) for any cancer and 71 years (SD = 12) for colorectal cancer. The overall mean age at enrollment was 49 years, which

Figure 1. Associations Between Adherence to WCRF/AICR[a] Recommendations and Cancer Risk



* Statistically significant.

[a] WCRF/AICR, World Cancer Research Fund/American Institute for Cancer Research.

[b] Cox proportional hazard ratio (HR) models were used to estimate hazard ratios and 95% CIs (as the horizontal whiskers show) for the association of meeting WCRF/AICR cancer prevention recommendations with cancer risk. Hazard Ratio (HR) models adjusted for age (continuous, years; age at interview), sex (male/female), and smoking status (categorial, current/former/never) for both sexes combined cohort. We report the ratio between the participants who were most adherent to the recommendations (individual score: 1; composite score: 4) to individuals not meeting WCRF/AICR recommendations (individual /composite score: 0) for each score category.

[c] CIs are very wide, which indicates that we have little knowledge about the effect and further information is needed.

was statistically significantly higher in cases (63 years) compared to controls (48 years).

In the overall cohort, the distribution of most sociodemographic characteristics at baseline was statistically significantly different ($P < .05$ in 2-sided χ^2 tests for categorical variables or Fisher tests for continuous variables) between cases and controls. Generally, cases had a higher proportion of lower education, lower income, white race, widowed/separated or divorced, newly immigrated, living in rural regions, and smoker participants. In the colorectal cancer cohort, the distribution pattern of baseline sociodemographic variables was very similar to the overall cohort, while immigrant status was not associated with colorectal cancer case status ($P = .24$).

Proportions of participants meeting selected WCRF/AICR recommendations by sex and cancer cohort are presented in Table 2. Overall, except for adherence to physical activity recommendations, females reported higher adherence scores than males to the individual WCRF/AICR health recommendations in either the overall cohort or colorectal cancer cohort.

The proportion of participants meeting selected WCRF/AICR recommendations by case and control status in the any cancer and colorectal cancer cohorts is presented in Table 4. In the overall cohort, the distribution of the individual and composite adherence scores were all statistically significantly different ($P < .05$) between cases and controls. In the colorectal cancer cohort, except for the plant foods intake score ($P = .11$) and alcoholic drinking score ($P = .07$), for which the adherence scores were similar between cases and controls, the other 3 adherence score variables (body fatness score, physical activity score, and the composite adherence score) were all statistically significantly different between cases and controls ($P < .0001$). Generally, a greater proportion of controls than cases followed the recommendations.

The associations between adherence to individual WCRF/AICR recommendations and the risk of developing any cancer or colorectal cancer are presented in Figure 1 and in Table 5. In the overall cohort, lower body fatness (HR, 0.90; 95% CI, 0.86–0.94), decreasing alcoholic drinks (HR, 0.88; 95% CI, 0.78–0.99), and increasing vegetable and fruit consumption (HR, 0.90; 95% CI, 0.84–0.97) each were associated with reduced risk of developing cancer. However, the findings were not statistically significant for increasing physical activity. Similar findings were observed for the colorectal cancer cohort, except for increasing plant food consumption, which was not associated with a decreased risk of developing colorectal cancer.

When the overall cohort was stratified by sex, the statistically significant findings for males were the same as those observed for both sexes combined. For females, however, only lower body fatness (HR, 0.89; 95% CI, 0.84–0.95) and increased physical activity (HR, 0.90; 95% CI, 0.83–0.98) were associated with a decreased risk of any cancer. When the colorectal cancer cohort was stratified by sex, each individual recommendation along with the composite score were associated with a decreased risk of colorectal cancer for males, but only lower body fatness (HR, 0.76; 95% CI, 0.64–0.91) was associated with decreased colorectal cancer risk for females.

HRs for the risk of developing any cancer and colorectal cancer with the composite adherence score are also presented in Figure 1. After adjusting for confounding factors (age, sex, and smoking status), compared to individuals not meeting any WCRF/AICR recommendations (composite score, 0), participants who were most adherent to the recommendations (composite score, 4) were 31% less likely to develop any cancer (HR, 0.69; 95% CI, 0.51–0.92), and were 61% less likely to develop colorectal cancer (HR, 0.39; 95% CI, 0.20–0.77).

When stratified by sex, the negative associations remained statistically significant for both male cohorts. Compared to male individuals not meeting WCRF/AICR recommendations, the most adherent male group was 40% less likely to develop any cancer (HR, 0.60; 95% CI, 0.43–0.86) and 73% less likely to develop colorectal cancer (HR, 0.27; 95% CI, 0.12–0.63). The findings were less consistent among females, as the association was statistically significant for the any cancer female cohort ($P = .0001$), but not statistically significant for the colorectal cancer female cohort ($P = .3214$).

For all statistically significant adherence scores, the magnitude of reduced risk increased with increasing adherence to the recommendations when examining partial adherence to a recommendation or composite adherence score (Table 5).

Discussion

This study examined whether following 4 selected WCRF/AICR cancer prevention recommendations was associated with lower cancer risk in Ontario. Overall, the findings from our study support the WCRF/AICR cancer prevention recommendations and suggests that a large proportion of cancer cases may be prevented through management of modifiable risk factors in Ontario. The WCRF/AICR report presented 8 general cancer prevention recommendations, and we were able to examine the 4 factors that were reported in the CCHS data for Ontario respondents; specifically, body fatness, physical activity, plant foods, and alcoholic drinks. Data on the other 4 WCRF/AICR recommendations on dietary components (foods and drinks that promote weight gain; animal foods; dietary supplements use; and food preservation, processing, and preparation) were not available and could not be evaluated in our study.

Consistent with our findings, previous large cohort studies reported that a stronger adherence to WCRF/AICR cancer prevention recommendations was associated with reduced risk of developing any cancer. A study from Alberta, Canada found participants who were the most adherent to the cancer prevention recommendations were 13% less likely to develop cancer compared to the least adherent group.[9] A study from France also found that following WCRF/AICR recommendations was associated with a statistically significantly reduced risk of overall cancer by 12%.[8] A recent meta-analysis of 38 studies found adhering to the WCRF/AICR recommendations was associated with lower risks of cancer incidence, especially for breast and colorectal cancer.[24] This is consistent with our

findings for all cancers and for colorectal cancer, suggesting that cancer recommendations are most associated with cancer incidence when the recommendations are based on modifiable risk factors such as diet.

In a cohort study from 9 European countries, the authors found that adherence to the WCRF/AICR recommendations was associated with decreased risk of any cancer and for specific cancer sites including breast, endometrial, colorectal, lung, kidney, stomach, upper aerodigestive tract, liver, and esophageal.[11] Another cohort study of older adults (ages ≥60 years) in European countries and the United States found a lower risk of total cancer, colorectal cancer, and prostate cancer with each 1-point increase in an individual's WCRF/AICR diet score.[10] This is consistent with our study's findings, as the magnitude of reduced risk increased with increasing adherence to the recommendations for the statistically significant adherence score. Findings from our study, as well as other studies in Canada, the United States, and Europe, suggest that the WCRF/AICR cancer prevention recommendations provide valuable guidance to reduce the risk of developing any cancer and many specific cancers, including colorectal cancer.

In our study, following all 4 selected WCRF/AICR cancer prevention recommendations was associated with lower risk of developing any cancer and colorectal cancer in both sexes combined and among males only, but no statistically significant associations were observed among females. The findings from a study in Alberta, Canada were not consistent with our study when stratified by sex, as the associations for women remained statistically significant but not for men in Alberta.[9] This may be due to differences in the operationalization of the WCRF/AICR recommendations, as the Alberta study examined 6 recommendations with binary adherence (1 or 0) as opposed to our study, which assessed 4 recommendations with options for partial adherence (1.0, 0.5, 0.0).[9] Further research to further explore possible differences by sex is warranted. Also, the interaction term between the potential effect modifier of interest and the continuous lifestyle score was not introduced into this study but is a potential area of future research.

Our study is the first comprehensive linkage population-based study of cancer incidence associated with modifiable risk factors in Ontario, and there are some limitations to note. As noted earlier, we constructed the composite adherence score by assigning equal weight to the 4 individual risk factors (body fatness, physical activity, plant foods, and alcoholic drinks). Since these 4 individual factors have different magnitudes of association with cancer risk, weighting variables accordingly in the composite adherence score may improve the model efficiency and accuracy.[9] Survey error and reporting bias may also exist in the CCHS responses because it is based on self-reported body fatness, diet, and physical activity, so potential underreporting on unhealthy behaviors may affect the accuracy and reliability of some results.[25]

Some other potential confounders such as comorbidities, family history of cancer, and exposure to other risk factors may impact an individual's risk of developing cancer, but these were beyond the scope of this study. The study also does not consider that cancer is likely caused by the interplay of many risk factors and therefore it is important to assess the integrated effect of multiple cancer risk factors.[26]

Furthermore, data on CCHS participants' health and lifestyle behaviors were only captured at 1 point in time based on their behaviors during the 12 months prior to survey enrollment. Therefore, we were unable to capture any behavior changes over time in our analysis. At the time of data linkage, data for follow up was only available until 2014; however, additional years of follow up could be included in future research. A final limitation to consider is that our operationalization of the WCRF/AICR recommendations was not based on validated measures and was limited by the alignment of CCHS survey questions with the recommendations as defined by the WCRF/AICR.

Conclusion

Ontarians who adhered to the WCRF/AICR cancer prevention recommendations had a lower risk of developing any cancer and colorectal cancer compared to individuals not meeting WCRF/AICR recommendations, and the magnitude of reduced risk increased with increasing adherence to the recommendations. Having a BMI within the normal range, increasing vegetable and fruit consumption, increasing physical activity, and decreasing alcohol consumption were associated with a decreased risk of developing any cancer (31%) or developing colorectal cancer (61%). When stratified by sex, the associations remained statistically significant for men. Overall, the findings from our study contribute to the growing body of evidence indicating that healthy behaviors consistent with selected WCRF/AICR recommendations are associated with a decreased risk of developing any cancer and colorectal cancer.

The risk of developing cancer can be significantly reduced by adopting healthy behaviors and the WCRF/AICR cancer prevention recommendations provide guidance in doing so. Following these recommendations can also reduce the risk of other chronic diseases. In Ontario, resources such as the My CancerIQ risk assessment tool feature such messaging.[27] This study highlights the value of translating etiologic research on cancer risk factors into recommendations for the general public, as well as the importance of population health approaches to decrease cancer risk factor exposure and reduce the burden of cancer in the province.

Acknowledgements

We would like to sincerely thank Elisa Candido, Shelley Harris, Julie Klein-Geltink, and Alice Peter for their contributions to the conceptualization of the research question and initial efforts to acquire data.

References

1. CCO and Ontario Agency for Health Protection and Promotion (Public Health Ontario). *The Burden of Chronic Diseases in Ontario: Key Estimates to Support Efforts in Prevention*. Queen's Printer for Ontario; 2019.
2. Cancer Care Ontario. *Ontario Cancer Statistics 2018*. Cancer Care Ontario; 2018.

3. Poirier AE, Ruan Y, Volesky KD, et al. The current and future burden of cancer attributable to modifiable risk factors in Canada: summary of results. *Prev Med*. 2019;122:140-147.
4. Vineis P, Alavanja M, Buffler P, et al. Tobacco and cancer: recent epidemiological evidence. *J Nat Cancer Inst*. 2004;96(2):99-106.
5. Pischon T, Nöthlings U, Boeing H. Obesity and cancer. *Proc Nutr Soc*. 2008;67(2):128-145.
6. Friedenreich CM, Orenstein MR. Physical activity and cancer prevention: etiologic evidence and biological mechanisms. *J Nutr*. 2002;132(11 suppl):3456s-3464s.
7. World Cancer Research Fund/American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective*. American Institute for Cancer Research; 2007.
8. Lavalette C, Adjibade M, Srour B, et al. Cancer-specific and general nutritional scores and cancer risk: results from the prospective NutriNet-Sante cohort. *Cancer Res*. 2018;78(15):4427-4435.
9. Xu JY, Vena JE, Whelan HK, Robson PJ. Impact of adherence to cancer-specific prevention recommendations on subsequent risk of cancer in participants in Alberta's Tomorrow Project. *Public Health Nutr*. 2019;22(2):235-245.
10. Jankovic N, Geelen A, Winkels RM, et al. Adherence to the WCRF/AICR dietary recommendations for cancer prevention and risk of cancer in elderly from Europe and the United States: a meta-analysis within the CHANCES project. *Cancer Epidemiol Biomarkers Prev*. 2017;26(1):136-144.
11. Romaguera D, Vergnaud AC, Peeters PH, et al. Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study. *Am J Clin Nutr*. 2012;96(1):150-163.
12. Petimar J, Smith-Warner SA, Rosner BA, Chan AT, Giovannucci EL, Tabung FK. Adherence to the World Cancer Research Fund/American Institute for Cancer Research 2018 recommendations for cancer prevention and risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2019;28(9):1469-1479.
13. Romaguera D, Gracia-Lavedan E, Molinuevo A, et al. Adherence to nutrition-based cancer prevention guidelines and breast, prostate and colorectal cancer risk in the MCC-Spain case-control study. *Int J Cancer*. 2017;141(1):83-93.
14. Turati F, Bravi F, Di Maso M, et al. Adherence to the World Cancer Research Fund/American Institute for Cancer Research recommendations and colorectal cancer risk. *Eur J Cancer*. 2017;85:86-94.
15. Hastert TA, White E. Association between meeting the WCRF/AICR cancer prevention recommendations and colorectal cancer incidence: results from the VITAL cohort. *Cancer Causes Control*. 2016;27(11):1347-1359.
16. Statistics Canada. Canadian Community Health Survey (CCHS) Annual Component. *User Guide: 2007-2008 Microdata File*. Statistics Canada; 2009.
17. Clarke EA, Marrett LD, Kreiger NR. *Twenty Years of Cancer Incidence 1964-1983: The Ontario Cancer Registry*. Ontario Cancer Treatment and Research Foundation; 1987.
18. National Cancer Institute. Surveillance, Epidemiology and End Results Program ICD-O-3 Site Recode 2008. **https://seer.cancer.gov/siterecode**
19. SAS/ACCESS 9.4 Interface to ADABAS: Reference. SAS Institute Inc; 2013.
20. Dorak MT, Karpuzoglu E. Gender differences in cancer susceptibility: an inadequately addressed issue. *Front Genet*. 2012;3:268.
21. Kim HI, Lim H, Moon A. Sex differences in cancer: epidemiology, genetics and therapy. *Biomol Ther*. 2018;26(4):335-342.
22. Jacob L, Freyn M, Kalder M, Dinas K, Kostev K. Impact of tobacco smoking on the risk of developing 25 different cancers in the UK: a retrospective study of 422,010 patients followed for up to 30 years. *Oncotarget*. 2018;9(25):17420-17429.
23. US Cancer Statistics Working Group. US cancer statistics: 1999–2009 incidence and mortality web-based report. Centers for Disease Control and Prevention website. **https://www.cdc.gov/uscs**
24. Solans M, Chan DSM, Mitrou P, Norat T, Romaguera D. A systematic review and meta-analysis of the 2007 WCRF/AICR score in relation to cancer-related health outcomes. *Ann Oncol*. 2020;31(3):352-368.
25. Shields M, Connor Gorber S, Janssen I, Tremblay MS. Bias in self-reported estimates of obesity in Canadian health surveys: an update on correction equations for adults. *Health Rep*. 2011;22(3):35-45.
26. Emmons KM. Maximizing cancer risk reduction efforts: addressing multiple risk factors simultaneously. *Cancer Causes Control*. 1997;8(suppl 1):S31-S34.
27. My CancerIQ. Cancer Care Ontario website. Updated October 21, 2021. **https://mycanceriq.ca/**

Original Article

Cancer among Refugees Resettled to Idaho during 2008–2019: A Proof-of-Concept Study

Bożena M. Morawski, PhD, MPH[a]; Randi Rycroft, MSPH, CTR[a]; Christopher J. Johnson, MPH[a]

Abstract: Background: Disparities in cancer burden and outcomes according to socioeconomic characteristics have been extensively characterized for US populations. The cancer experience of refugees, who may share characteristics of other socioeconomically disadvantaged populations and also experience distinct barriers to care, has not been described previously. We conducted a proof-of-concept study evaluating our ability to characterize cancer incidence in refugees resettled to Idaho via a novel linkage of cancer data and administrative data characterizing refugee arrivals to Idaho. Methods: In July 2021, the Cancer Data Registry of Idaho probabilistically linked cancer surveillance data and refugee arrival data (2008–2019 diagnosis and arrival years) collected through the Centers for Disease Control and Prevention's Electronic Disease Notification (EDN) System. We used SEER*Stat to calculate standardized incidence ratios (SIR) for malignant tumors and benign/borderline malignant brain and other nervous system (ONS) tumors using Idaho-specific and Surveillance, Epidemiology, and End Results (SEER) Program referent incidence rates. Results: 60 malignant and 7 benign brain and ONS tumors were diagnosed among 9,499 refugees resettled to Idaho. Refugees had fewer than expected malignant tumors overall (57 observed vs 96.0 expected; SIR, 0.60; 95% CI, 0.45–0.77). An excess of tumors of the esophagus were diagnosed among Southeast Asian refugees (4 observed vs 0.64 expected; SIR, 6.3; 95% CI, 1.7–16.0). We also used EDN data to update country of birth for linked persons. Conclusions: Linking EDN refugee data to cancer surveillance data presented unique challenges. However, we used a novel data source to augment cancer data and characterize incidence in refugees, potentially improving our ability to serve this vulnerable population.

Key words: *cancer, cancer registry, electronic disease notification system, health disparities, refugees*

Introduction

Refugees are persons who are unable or unwilling to return to their country of nationality "because of persecution or a well-founded fear of persecution due to race, religion, nationality, membership in a particular social group, or political opinion" (US Immigration and Nationality Act, Sect. 101[a][42]).[1] Each year, Idaho welcomes refugees from a diverse set of home countries, including, but not limited to, Afghanistan, Democratic Republic of Congo, Iraq, Bhutan, and Syria. A median of 816 refugees were resettled to Idaho per year during 2008–2019.

Some characteristics of refugee populations and other immigrant populations may be shared. For example, both groups may be comprised of minority racial/ethnic groups, have relatively lower income, and experience language- or transportation-related barriers in accessing health care. However, refugees also have legal immigration status and access to government-sponsored resources; in particular, health insurance. Refugee populations in Idaho have demonstrated a health profile distinct from other Idahoans, including an elevated prevalence of chronic medical conditions (eg, diabetes[2]) and psychological conditions (eg, posttraumatic stress disorder[3]). Cancer burden, however, has not been characterized at the population level among Idaho refugees.

Outside of the United States, aspects of the cancer experience among refugees have been described[4-8] and European cohorts are being established to more completely characterize the relationship between long-distance migration, noncommunicable diseases, and mental health among refugees and asylum seekers.[9] Much of the literature describing the cancer experience of refugees resettled to the United States focuses on screening behaviors,[10-12] which does not help clinical or public health practitioners understand the scope of the cancer burden among Idaho's refugees and potential disparities in incidence and outcomes.[13]

The Center for Disease Control and Prevention's Electronic Disease Notification (EDN) System, described in detail elsewhere,[1] has collected health, demographic, and administrative immigration data on legal permanent immigrants, refugees, asylees, and parolees who have entered the United States since 2006. The EDN System provides a centralized mechanism by which to notify state and local public health agencies of refugee arrivals, ensuring that arrivals meet immigration requirements and benefit from services provided on arrival, such as health screenings. In

[a]Cancer Data Registry of Idaho, Idaho Hospital Association, Boise, Idaho.

Address correspondence to Bożena M. Morawski, PhD, MPH, Cancer Data Registry of Idaho, 373 West Fort Street, Boise, ID 83701-1278. Telephone: (208) 489-1373. Fax: (208) 344-0180. Email: bmorawski@teamiha.org.

The Idaho Hospital Association (IHA) contracts with, and receives funding from, the Idaho Department of Health and Welfare, Division of Public Health, to provide a statewide cancer surveillance system: the Cancer Data Registry of Idaho (CDRI). This report has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN261201800006I and the Centers for Disease Control and Prevention, Department of Health and Human Services, under Cooperative Agreement 1NU58DP006270.

Idaho, the Idaho Department of Health and Welfare receives data on refugee arrivals via EDN for the entire state.

The Cancer Data Registry of Idaho (CDRI) is the population-based cancer registry for the state of Idaho, and collects incidence and survival data on all cancer patients who reside in Idaho and any non–Idaho residents diagnosed with or treated for cancer in Idaho. CDRI has been population-based since 1971. To our knowledge, the cancer experience for refugees has not been evaluated at the population level in Idaho or elsewhere in the United States. This study was conducted to address this gap in the literature and demonstrate proof of concept for linking data from EDN with cancer surveillance data to identify refugees resettled to Idaho and characterize cancer incidence in this population.

Methods

A file containing data for all refugees and special immigrant visa recipients who were resettled to Idaho from January 1, 2006, through December 31, 2020, was provided to the Cancer Data Registry of Idaho by the Idaho Department of Health and Welfare. The refugee cohort data set included date of arrival to the United States; first, last, and middle name; unique 9-digit alien identification number (with digits 1–6 being assigned to a family, and 7–9 assigned to an individual); date of birth; birth city and county; sex; country of residence at time of immigration; and Idaho jurisdiction of initial settlement. For refugees who were not initially settled to Idaho or migrated out of Idaho, and who provided formal notification of migration within the United States, the data set also includes date of arrival or departure from Idaho and the jurisdiction that the person moved from/to. Because Social Security numbers (SSNs) are not included in the EDN system, this identifier was not available. Additionally, a large proportion of refugee dates of birth are nonspecific (eg, January 1 of a given year).

Tumors that are reportable to either the state of Idaho, the National Program of Cancer Registries, or the SEER Program for the period from January 2008 through December 2019 with malignant behavior (International Classification of Diseases for Oncology, 3rd Edition [ICD-O-3] behavior recode for analysis) or benign or borderline malignant tumors of the brain and other nervous system were exported from SEER*DMS, CDRI's surveillance database. Tumors with unknown year of diagnosis were excluded.

A probabilistic linkage was performed between EDN and CDRI data using Match*Pro version 1.6.5 (released April 18, 2021; Information Management Services, Inc). Candidate matches were identified using last name, first name, middle name, date of birth, and county of residence at time of cancer diagnosis and county of resettlement. The number of potential match pairs was reduced by blocking on SOUNDEX functions of the name fields and year or month and day of date of birth. Dates of diagnosis and arrival to the United States were used to evaluate potential matches. Manual review was conducted on all candidate matches using supporting information (eg, immigration date versus date of diagnosis) and additional resources (eg, LexisNexis). Records were evaluated at the person level, and all qualifying cancer cases matched to a person in the refugee data set were included in the analytic linked data sets.

We created 2 analytic data sets, 1 for malignant tumors and 1 for benign and borderline malignant tumors of the brain and other nervous system. The analytic data sets included 1 record for each refugee-linked tumor and 1 record for each nonlinked refugee (ie, refugees who were not identified as being treated for or diagnosed with a cancer while an Idaho resident). Refugees arriving in the United States and tumors diagnosed before January 2008 and after December 2019 were excluded. Analytic data sets were prepared using SAS version 9.4 (SAS Institute Inc) and the SEER Windows FixLen Executable (updated November 2020; Information Management Systems, Inc).

Because no population denominators are available for the number of refugees living in Idaho at any given time, cancer incidence among refugees was calculated using person-time of residency in Idaho (date of immigration to the United States through the end of the follow-up period, December 31, 2019, or date of death or other qualifying event). The expected number of incident cancer cases was calculated using age-, sex-, and time period–specific referent cancer incidence rates. Because referent rates were not yet available for 2019, 2015–2018 incidence rates were applied to 2019 follow-up time. We compared the number of tumors diagnosed among refugees to the number of tumors that would be expected to be diagnosed among refugees, provided they had similar rates of cancer as in referent rate data. Referent rates from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program[14] and Idaho incidence data for all races combined were used to calculate the numbers of expected cases. We used the Multiple Primary-Standardized Incidence Ratio (MP-SIR) session in SEER*Stat version 8.3.9.1 (Information Management Systems, Inc)[15] to estimate standardized incidence ratios (SIRs) and corresponding 95% CIs for 46 primary site categories, which included 43 malignant tumor categories and 3 categories of benign brain and other nervous system tumors. Data sets were prepared for SEER*Stat in SEER*Prep version 2.6.0 (updated March 10, 2021; Information Management Systems, Inc) using a custom .dd file created for generalized applications of the MP-SIR session, and used previously by CDRI in cancer cluster investigations.[16] Tumors were classified based on histology and behavior according to the ICD-O-3/World Health Organization 2008 SEER Site Recode classification system.[17]

Analyses were conducted using 3 different latency periods: no latency period, 2 months, and 6 months. Because of the assumed long latency period for most cancers, we assumed that potential exposures and associated tumorigenesis likely began prior to arrival in the United States, especially for those tumors diagnosed within a year of arrival. However, as these analyses do not attempt to identify cancers associated with a particular exposure (eg, workplace-associated cancers), and because all cancers

Table 1. Demographic Information for Refugees Resettled to Idaho during 2008–2019 and Diagnosed with Cancer during 2008–2019, Compared to other Idahoans Diagnosed with ≥1 Tumor during 2008–2019

	Idaho refugee cancer survivors	*All other Idaho cancer survivors*
	(n = 65), n (%) or median [IQR]	**(n = 99,201)*, n (%) or median [IQR]**
Female, %	39 (60.0)	48,777 (49.2)
Age, years	52 [41, 69]	66 [57, 75]
Race**		
White	25 (38.5)	96,108 (96.9)
Black	10 (15.4)	325 (0.3)
Korean, Thai, Asian Indian	7 (10.8)	117 (0.1)
Other Asian	19 (29.2)	265 (0.3)
Other	4 (6.2)	283 (0.3)
Unknown	0 (0.0)	982 (1.0)

*First of any tumors diagnosed during 2008–2019.
**With the exception of unknown, comparison race categories selected per the distribution of Idaho refugee cancer survivors. All races for Idaho cancer survivors not presented.

diagnosed in this population are potentially informative to clinical and public health professionals, an analysis with no latency period is appropriate and perhaps the most informative. Estimates based on fewer than 4 tumors were suppressed.

Results

There were 9,499 refugees resettled to Idaho during 2008–2019, with a median of 816 persons resettled per year (minimum of 421 in 2017 and maximum of 1,221 in 2009). Approximately 48% (4,533/9,499) of refugees were female. The median age of refugees at the time of resettlement to the United States was 22.0 years (interquartile range [IQR], 10.0–34.0), while the median age of Idaho residents was 36.9 years in 2019. Although refugees resettled to Idaho were born in more than 80 countries, nearly 70% of refugees resettled to Idaho were born in 8 countries: Iraq (n = 1,374), Democratic Republic of Congo (n = 1,371), Bhutan (n = 1,094), Myanmar (n = 937), Tanzania (n = 500), Ethiopia (n = 422), Nepal (n = 412), and Afghanistan (n = 400). January 1 of a given year was listed as the date of birth for 23.4% of refugees resettled to Idaho during this time period (2,227/9,499). Race and ethnicity information is not available for the refugee cohort.

Linkage results yielded 69 tumors diagnosed during 2008–2019, including 60 malignant cancers, 2 in situ, and 7 benign and borderline malignant behavior neoplasms of the brain and other nervous system. The mean ages at time of cancer diagnosis among refugees for all malignant primary sites and benign and borderline malignant tumors were 54.3 years and 55.9 years, respectively. In 2019, the mean age at diagnosis among Idaho residents diagnosed with a malignant tumor was 66.1 years, and 61.6 years for Idaho residents diagnosed with a benign or borderline malignant tumor. Demographic information for refugees diagnosed with cancer in Idaho is described in Table 1.

Three malignant tumors were excluded from analyses because they were diagnosed prior to or within 1 month of arrival in the United States, equating to 0 months of follow-up time in the MP-SIR session; a total of 57 malignant tumors and 7 benign and borderline malignant tumors were included in SIR analyses. Mean time from arrival to the end of the follow-up period was 6.47 years. The median number of years from arrival to diagnosis was 2 (IQR, 1–5 years; maximum time to diagnosis, 10 years). Approximately 44% (25/57) of tumors diagnosed among refugees were late stage cancers (distant or regional stage), versus the 41.5% of tumors diagnosed at late stage in 2018 among all Idaho residents; this difference was not statistically significant.[18]

Overall, for all malignant cancer sites combined, refugees had a lower-than-expected number of cancer diagnoses (57 observed vs 96.0 expected), with an observed-to-expected (O/E) ratio of 0.60. This difference was statistically significant, with a 95% CI for the O/E ratio that did not include 1 (0.45–0.77). Under conditions of no lag time, cases of breast cancer and cancers of the male genital system (driven by a lower-than-expected number of cancers of the prostate) were also statistically significantly lower among refugees than in the SEER-18 population.

For most other comparisons, the number of cancer diagnoses in refugees was statistically equivalent to the SEER-18 jurisdictions and Idaho, with the exception of cancers of the esophagus. The number of cancers of the esophagus diagnosed among refugees was statistically significantly higher than expected based on the SEER-18 referent population (SEER-18 O/E ratio = 6.3; 95% CI, 1.7–16.0; n = 4). These cancers were all diagnosed among refugees born in Southeast Asian countries. Results were consistent across lag periods of 0, 2, and 6 months. Results from SIR comparisons for malignant tumors are shown in Tables 2 and 3, using SEER-18 and Idaho referent rates.

The number of cases of benign and borderline malignant tumors of the brain and other central nervous system diagnosed in Idaho's refugees was similar to the expected number, based on SEER-18 and Idaho referent rates. Results from SIR comparisons for benign and borderline malignancies of the brain and other nervous system tumors are shown in Tables 4 and 5, using SEER-18 and Idaho referent rates.

Discussion

This proof-of-concept study showed that linkage between refugee and cancer registry data sets was feasible, and that estimates of cancer burden for the refugee population may be distinct from nonrefugee Idahoans. For Idaho, linkage between the refugee data from EDN and cancer surveillance data yielded high certainty matches, despite incomplete dates of birth and the absence of SSNs. CDRI was also able to update registry data on country of birth for linked patients, which improved specificity of race classification. Establishing reliable results of the linkage for other states that have larger and more ethnically diverse populations may prove more challenging.

Table 2. Malignant Tumors Diagnosed during 2008–2019 among Refugees Resettled to Idaho during 2008–2019, as Classified by the SEER Site Recode ICD-O-3/WHO 2008, and Compared to SEER-18 (November 2020 Submission) Referent Rates

	Observed	*Expected*	*O/E*	*95% CI*	*Mean age at event, y*
All sites	57	95.97	0.59*	0.45–0.77	54.47
Digestive system	15	19.92	0.75	0.42–1.24	62.43
Esophagus	4	0.64	6.25*	1.7–16	64.22
Stomach	4	2.35	1.7	0.46–4.36	72.76
Colon and rectum	5	9.38	0.53	0.17–1.24	57.7
Respiratory system	9	9.47	0.95	0.43–1.80	68.25
Lung and bronchus	7	8.73	0.8	0.32–1.65	66.81
Male genital system	4	11.20	0.36*	0.1–0.91	56.55
Prostate	4	10.18	0.39	0.11–1.01	56.55
Urinary system	7	5.68	1.23	0.5–2.54	56.13
Kidney and renal pelvis	4	3.40	1.18	0.32–3.01	48.35
Endocrine system	9	5.72	1.57	0.72–2.99	43.48
Thyroid	9	5.37	1.68	0.77–3.18	43.48
Leukemia	4	3.10	1.29	0.35–3.3	44.58

ICD-O-3, International Classification of Diseases for Oncology, 3rd Edition; O/E, observed-to-expected ratio; SEER, Surveillance, Epidemiology, and End Results Program; WHO, World Health Organization.

Site category rows with small cell suppression (< 4 cases) are excluded for succinctness, including breast cancer. *$P < .05$.

Table 3. Malignant Tumors Diagnosed during 2008–2019 among Refugees Resettled to Idaho during 2008–2019, as Classified by the SEER Site Recode ICD-O-3/WHO 2008, and Compared to Idaho-Specific (November 2020 Submission) Referent Rates

	Observed	*Expected*	*O/E*	*95% CI*	*Mean age at event, y*
All Sites	57	86.98	0.66*	0.5–0.85	54.47
Digestive system	15	17.98	0.83	0.47–1.38	62.43
Esophagus	4	1.21	3.31	0.9–8.48	64.22
Stomach	4	2.25	1.78	0.48–4.55	72.76
Colon and rectum	5	7.51	0.67	0.22–1.55	57.7
Respiratory system	9	9.38	0.96	0.44–1.82	68.25
Lung and bronchus	7	8.78	0.8	0.32–1.64	66.81
Male genital system	4	10.33	0.39*	0.11–0.99	56.55
Prostate	4	9.64	0.41	0.11–1.06	56.55
Urinary system	7	7.07	0.99	0.4–2.04	56.13
Kidney and renal pelvis	4	3.22	1.24	0.34–3.18	48.35
Endocrine system	9	6.58	1.37	0.63–2.6	43.48
Thyroid	9	6.04	1.49	0.68–2.83	43.48
Leukemia	4	2.91	1.37	0.37–3.52	44.58

ICD-O-3, International Classification of Diseases for Oncology, 3rd Edition; O/E, observed-to-expected ratio; SEER, Surveillance, Epidemiology, and End Results Program; WHO, World Health Organization.

Site category rows with small cell suppression (< 4 cases) are excluded for succinctness, including breast cancer. *$P < .05$.

Table 4. Benign and Borderline Malignant Tumors Diagnosed during 2008–2019 among Refugees Resettled to Idaho during 2008–2019, as Classified by the SEER Site Recode ICD-O-3/WHO 2008, and Compared to SEER-18 (November 2020 Submission) Referent Rates

	Observed	*Expected*	*O/E*	*95% CI*	*Mean age at event, y*
Brain and other nervous system	7	3.50	2	0.8–4.12	55.94

ICD-O-3, International Classification of Diseases for Oncology, 3rd Edition; O/E, observed-to-expected ratio; SEER, Surveillance, Epidemiology, and End Results Program; WHO, World Health Organization.

Site category rows with small cell suppression (< 4 cases) are excluded for succinctness.

Table 5. Benign and Borderline Malignant Tumors Diagnosed during 2008–2019 among Refugees Resettled to Idaho during 2008–2019, as Classified by the SEER Site Recode ICD-O-3/WHO 2008, and Compared to Idaho-Specific (Nov 2020 Submission) Referent Rates

	Observed	*Expected*	*O/E*	*95% CI*	*Mean age at event, y*
Brain and other nervous system	7	4.04	1.73	0.7–3.57	55.94

ICD-O-3, International Classification of Diseases for Oncology, 3rd Edition; O/E, observed-to-expected ratio; SEER, Surveillance, Epidemiology, and End Results Program; WHO, World Health Organization.

Site category rows with small cell suppression (< 4 cases) are excluded for succinctness.

From a cancer burden standpoint, these data demonstrate that overall, refugees have lower rates of cancer incidence than other Idaho residents, even when accounting for differences in age distributions of the refugee and nonrefugee populations. In instances where lower-than-expected SIRs were for screening-related cancers (eg, breast and prostate; data suppressed per suppression rules outlined in methods), further investigation is warranted to determine if screening outreach should be increased in the refugee population. We also identified certain cancers that were more common in refugees than the general Idaho population, specifically esophageal cancers among refugees born in Southeast Asia. Human papilloma virus (HPV) in particular has been linked to esophageal cancers in the Asian region, with HPV present in up to 46.5% of tumors.[19] Heavy alcohol use in combination with smoking is another major contributing cause of esophageal cancer. Further investigation is warranted to determine the root causes of these differences in observed versus expected cancers and what, if any, clinical and public health interventions are appropriate for Southeast Asian refugee populations.

These analyses had limitations that will be addressed in future analyses. First, linkage sensitivity and specificity might be improved by linking refugee data to commercially available credit products (eg, LexisNexis or Accurint) and determining SSN. Second, CDRI was unable to obtain refugee vital status via linkage with Idaho vital statistics data, meaning that follow-up time for refugees who died prior to the end of the study period but not linked to the CDRI data set as cancer survivors was inflated. This lack of vital statistics data would, however, attenuate results towards the null. Additionally, unpublished data evaluating mortality in this cohort in a similar period (follow-up through December 2017 for arrivals during 2011–2017) yielded only 33 deaths, indicating that the addition of vital status, while critical for statistically precise estimates of person-time, might not meaningfully impact results. Finally, tumor data for 2019 were incomplete, with the corresponding implication that 2019 estimates may change when data are finalized.

To our knowledge, this is the first time that cancer incidence has been specifically estimated for the unique population of refugees resettled to the United States. However, it is difficult to contextualize cancer incidence without corresponding rates of cancer-related mortality and cancer survival among refugees when compared to other Idahoans or other US residents.[20] These types of analyses could provide important information for clinical and public health stakeholders and provide an opportunity to improve country of birth information present in national surveillance data. Future planned analyses will include updated 2019 and 2020 cancer incidence estimates and estimates of cancer-related mortality and survival in the refugee population. In addition, the ability to partner with larger and more ethnically diverse states with more refugees and more tumors will allow the methods from this proof-of-concept study to be refined and improved so that they are appropriate for a larger number of US population-based cancer registries.

Acknowledgements

The statewide cancer registry database is a product of collaboration among many report sources, including hospitals, physicians, surgery centers, pathology laboratories, and other states in which Idaho residents are diagnosed or treated for cancer. Their cooperation in reporting timely, accurate, and complete cancer data is acknowledged and sincerely appreciated.

CDRI would also like to thank the Division of Public Health, Idaho Department of Health and Welfare for their continued partnership in cancer control and prevention.

CDRI would also like to thank Steve Scoppa of Information Management Services for the custom MP-SIR session used to create analytic data sets for this project, and for technical support in their implementation.

References

1. Lee D, Philen R, Wang Z, et al. Disease surveillance among newly arriving refugees and immigrants—electronic disease notification system, United States, 2009. *MMWR Surveill Summ*. 2013;62(7):1-20.
2. Smith M, Springer P, Soelberg T, Lazare P, Temkin-Martinez M. Health conditions of post-resettlement African refugees in Boise, Idaho. *Online J Cult Competence Nurs Healthcare*. 2016;6(1):70-82.
3. Begic S, McDonald TW. The psychological effects of exposure to wartime trauma in Bosnian residents and refugees: implications for treatment and service provision. *Int J Ment Health Addict*. 2006;4(4):319-329.
4. Thordardottir EB, Yin L, Hauksdottir A, et al. Mortality and major disease risk among migrants of the 1991-2001 Balkan wars to Sweden: a register-based cohort study. *PLoS Med*. 2020;17(12):e1003392.
5. Swerdlow A. Mortality and cancer incidence in Vietnamese refugees in England and Wales: a follow-up study. *Int J Epidemiol*. 1991;20(1):13-19.
6. Kebudi R, Bayram I, Yagci-Kupeli B, et al. Refugee children with cancer in Turkey. *Lancet Oncol*. 2016;17(7):865-867.
7. Grulich AE, Swerdlow AJ, Head J, Marmot MG. Cancer mortality in African and Caribbean migrants to England and Wales. *Br J Cancer*. 1992;66(5):905-911.
8. Norredam M, Olsbjerg M, Petersen JH, Juel K, Krasnik A. Inequalities in mortality among refugees and immigrants compared to native Danes—a historical prospective cohort study. *BMC Public Health*. 2012;12:757.
9. Eiset AH, Aoun MP, Haddad RS, et al. Asylum seekers' and refugees' changing health (ARCH) study protocol: an observational study in Lebanon and Denmark to assess health implications of long-distance migration on communicable and non-communicable diseases and mental health. *BMJ Open*. 2020;10(5):e034412.
10. Siddiq H, Alemi Q, Mentes J, Pavlish C, Lee E. Preventive cancer screening among resettled refugee women from Muslim-majority countries: a systematic review. *J Immigr Minor Health*. 2020;22(5):1067-1093.
11. Abdi HI, Hoover E, Fagan SE, Adsul P. Cervical cancer screening among immigrant and refugee women: scoping-review and directions for future research. *J Immigr Minor Health*. 2020;22(6):1304-1319.
12. Raines Milenkov A, Felini M, Baker E, et al. Uptake of cancer screenings among a multiethnic refugee population in North Texas, 2014-2018. *PLoS One*. 2020;15(3):e0230675.
13. 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.
14. Surveillance, Epidemiology, and End Results (SEER) Program (**www.seer.cancer.gov**) SEER*Stat Database: Incidence - SEER Research Data, 18 Registries, Nov 2020 Sub (1992–2018) - Linked To County Attributes - Time Dependent (1990-2018) Income/Rurality, 1969–2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission.
15. Surveillance Research Program NCI. SEER*Stat. 8.3.9.1 ed.
16. Rosenthal M, Johnson CJ, Scoppa S, Carter K. Two suspected worksite or occupational cancer clusters investigated using the cancer data registry and multiple primary standardized incidence ratios in SEER *Stat–Idaho, 2013–2014. *J Registry Manag*. 2016;41(3):128-133.
17. Site Recode ICD-O-3/WHO 2008 Definition. National Cancer Institute Surveillance, Epidemiology, and End Results Program website. **https://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html**
18. Johnson CJ, Morawski BM, Rycroft RK. *Cancer in Idaho – 2018*. Cancer Data Registry of Idaho; 2020.
19. Wang J, Zhao L, Yan H, et al. A meta-analysis and systematic review on the association between human papillomavirus (types 16 and 18) infection and esophageal cancer worldwide. *PLoS One*. 2016;11(7):e0159140.
20. Cho H, Mariotto AB, Schwartz LM, Luo J, Woloshin S. When do changes in cancer survival mean progress? The insight from population incidence and mortality. *J Natl Cancer Inst Monogr*. 2014;2014(49):187-197.

Original Article

Bias Introduced by Relying on Incomplete Electronic Pathology Reporting for Rapid Case Ascertainment in Patient Contact Studies

Margaret Gates Kuliszewski, ScD[a, b]; Jovanka N. Harrison, PhD[a]; Maria J. Schymura, PhD[a, b]

Abstract: Background: Relying on electronic pathology (ePath) reporting to state cancer registries for rapid ascertainment of cases for patient contact research studies may introduce bias if the patient populations differ for reporting facilities with vs without ePath. We examined changes between 2014–2019 in the percent of cases reported to the New York State Cancer Registry by ePath within 3 months of diagnosis and characteristics of the most recent cases by ePath status. Our goal was to assess the potential bias introduced by relying on incomplete ePath reporting for patient recruitment. Methods: We restricted our analysis to first malignant cancers diagnosed in New York State residents aged 18 years and older. We examined patient characteristics and used χ^2 tests to examine differences in the distribution of each characteristic by ePath status for diagnosis years 2017–2019, and used multivariable-adjusted logistic regression to calculate odds ratios and 95% CIs for the association between each patient characteristic and ePath status for all 2017–2019 cancers combined and common cancer sites. All analyses were conducted using SAS 9.4. Results: The percent of cases reported by ePath increased over time from 15.7% in 2014 to 44.8% in 2019. Among 264,607 cancers diagnosed in 2017–2019 and reported through July 2021, there were statistically significant differences in all variables examined by ePath status (all $P < .0001$). For all cancers combined, cases reported by ePath were more likely to be younger, female, non-Hispanic White, married, live outside of New York City/Long Island, still be alive, and have received treatment. We observed statistically significant odds ratios for the associations between all variables examined and ePath status for all cancers combined, but the strength and statistical significance of the associations varied by cancer site. Conclusions: Our results indicate that relying on incomplete ePath reporting for rapid case ascertainment will introduce selection bias in the study sample for patient contact studies. This bias should decrease as additional facilities acquire ePath reporting capability.

Key words: *cancer outcomes; electronic pathology reporting; New York State Cancer Registry; patient contact studies; rapid case ascertainment; selection bias; Surveillance, Epidemiology, and End Results (SEER) Program*

Introduction

State and regional cancer registries are a potential source of cases for research studies involving patient contact and seeking to answer questions related to cancer risk, treatment, and outcomes. Advantages of using cancer registries for research include the population-based nature of registries, where almost all cases occurring in a population are captured because of the thorough reporting and follow-up processes, and the detailed information collected for each case, including demographic, diagnostic, treatment, and survival information.[1] However, a possible limitation of using registries for patient contact studies is the time lag of 6 to 12 months or more between diagnosis of a case and availability of complete, consolidated case information, as researchers are often interested in contacting and enrolling cancer patients soon after diagnosis.

Electronic pathology (ePath) reports are a potential resource for earlier identification of newly diagnosed cases. These reports are routinely submitted to the New York State Cancer Registry (NYSCR) and other state and regional cancer registries by a subset of hospitals and other reporting facilities. Reports submitted by ePath include fewer data items and are more likely to have missing or incomplete data than full case reports submitted by hospital tumor registries.[2] However, ePath reports usually provide enough information to assess initial eligibility and contact a patient to invite them to participate in a study. A drawback of using ePath reports in research studies is the potential for differences in the patient populations for facilities with and without ePath reporting, which may introduce bias for studies that rely on ePath reports for rapid ascertainment of cases. To assess this potential bias, we examined changes over time in the percent of cases reported to the NYSCR by ePath and characteristics of the most recent cases by ePath status.

[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 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 opinions, findings, and conclusions expressed here are those of the authors and do not necessarily represent the official views of the funding agencies.

Methods

Study Population

Our study population included first malignant cancers in New York State residents aged 18 years and older who were diagnosed between 2014–2019 and reported to the NYSCR through July 2021. We retrieved data from the NYSCR Surveillance, Epidemiology, and End Results (SEER) Data Management System (SEER*DMS) database and excluded records with missing month of diagnosis or age at diagnosis and cases reported by nursing homes, autopsy, or death certificate only. For each consolidated case, we kept record-level information on the first ePath report received, if the case was reported by ePath, and otherwise the earliest case report received. After exclusions and restricting the data to 1 record per case, our study population included 258,285 cases diagnosed between 2014–2016, which were used for the assessment of changes in ePath reporting over time, and 264,607 cases diagnosed between 2017–2019, which were used for the assessment of changes in ePath reporting and differences in case characteristics by ePath status. Analyses of differences in case characteristics by ePath status focused on the most recent cases from 2017–2019 because these cases best reflect those that would be included in patient contact studies involving rapid case ascertainment.

Case Characteristics

A case was classified as an ePath report if SEER*DMS contained 1 or more ePath record reported to the NYSCR within 3 months after diagnosis. Cases without an ePath report and cases first reported by ePath more than 3 months after diagnosis were not considered ePath reports. We retrieved data from SEER*DMS on case characteristics of interest including age at diagnosis (categorized as 18–49, 50–59, 60–69, 70–79, or ≥80 years), sex (male, female, or other/unknown), race/ethnicity (non-Hispanic White, non-Hispanic Black, non-Hispanic Asian/Pacific Islander, Hispanic, or missing/other), marital status (single, married, divorced/separated, widowed, or other/unknown), region of residence (New York City [NYC]/Long Island, rest of state, or missing/unknown), vital status, stage at diagnosis (local, regional, distant, or unknown), and receipt of treatment during initial therapy (yes or no/unknown for chemotherapy, hormonal therapy, immunotherapy, radiation therapy, and surgery). Treatment was based on consolidated case information and therefore reflected all treatment received by the patient during initial therapy and reported to the NYSCR. Information on primary site was also retrieved and categorized using SEER site group codes as bladder (29010), breast (26000), colon/rectum (21041–21049, 21051–21052, or 21060), kidney (29020), leukemia (35011–35013, 35021–35023, 35031, 35041, or 35043), lung (22030), lymphoma (33011–33012 or 33041–33042), melanoma (25010), myeloma (34000), ovary (27040), pancreas (21100), prostate (28010), stomach (21020), thyroid (32010), uterus (27020), or other/miscellaneous (all other SEER site groups).

Statistical Analysis

We examined the percent of cancer cases reported by ePath by diagnosis year (for cases diagnosed in 2014–2019), as well as the overall percent of cases reported by ePath between 2017–2019. Subsequent analyses focused on the most recent cases diagnosed between 2017–2019, as these diagnosis years more accurately reflect current ePath reporting in the NYSCR. We examined descriptive characteristics of cases diagnosed in 2017–2019 by ePath status (categorized as "yes" if an ePath report was received within 3 months of diagnosis, and categorized as "no" otherwise), both for all 2017–2019 cases combined and by diagnosis year, and we calculated *P* values for the association with each descriptive characteristic using χ^2 tests. We repeated these analyses for each cancer site to assess how ePath reporting changed over time by primary site and to examine differences in case characteristics by ePath status for each primary site.

Finally, we used logistic regression to calculate multivariable-adjusted odds ratios (ORs) and 95% CIs for the association between each case characteristic of interest and ePath status for cases diagnosed in 2017–2019. For all cancers combined, logistic regression estimates were mutually adjusted for sex, age, race/ethnicity, marital status, region of residence, stage at diagnosis, treatment received during initial therapy, cancer site, and diagnosis year. For analyses of each cancer site, estimates were adjusted for all variables noted above except cancer site and, for sex-specific cancers, sex. All analyses were conducted using SAS version 9.4 (SAS Institute Inc).

Results

The percent of cases reported to the NYSCR by ePath within 3 months after diagnosis increased over time, with 15.7% reported by ePath in 2014, 27.5% in 2015, 31.5% in 2016, 34.6% in 2017, 35.5% in 2018, and 44.8% in 2019 (some results not shown). These increases over time resulted from additional facilities acquiring ePath reporting capability. The smaller increase in the percent of cases reported by ePath in 2018 compared to 2017 likely reflects reporting delays in 2018 that were related to implementation of new reporting guidelines.[3]

Of the 264,607 malignant first primary cancers diagnosed in New York State adults between 2017–2019, 101,474 were reported by ePath within 3 months after diagnosis (38.3%). The characteristics of these cases differed significantly from those not reported by ePath in terms of sex, age, race/ethnicity, marital status, region, vital status, stage at diagnosis, and receipt of treatment during initial therapy (all $P < .0001$; Table 1). Cases reported by ePath were more likely to be female (54.5% vs 48.0%), non-Hispanic White (69.3% vs 64.7%), married (53.8% vs 50.9%), local stage at diagnosis (49.4% vs 45.3%), and to have received treatment during initial therapy, including chemotherapy (32.0% vs 30.7%), hormonal therapy (23.7% vs 16.0%), immunotherapy (10.4% vs 8.6%), radiation therapy (32.1% vs 28.4%), and surgery (58.8% vs 48.6%), and were less likely to be 80 years of age or older (10.1% vs 14.8%), residents of NYC or Long Island (49.8% vs 53.9%), or deceased (16.3% vs 24.2%).

Table 1. Characteristics of 264,607 Malignant First Primary Cancers Diagnosed in New York State Residents Aged 18 Years and Older in 2017–2019 and Reported to the New York State Cancer Registry, by ePath Reporting Status*

	*Reported by ePath**		*P value†*
	No	**Yes**	
Total number (%)	163,133 (61.7)	101,474 (38.3)	
Case characteristics, n (%)			
Female	78,272 (48.0)	55,323 (54.5)	<.0001
Age category, y			<.0001
18–49	21,167 (13.0)	15,776 (15.6)	
50–59	30,391 (18.6)	21,971 (21.7)	
60–69	47,956 (29.4)	31,019 (30.6)	
70–79	39,463 (24.2)	22,447 (22.1)	
≥80	24,156 (14.8)	10,261 (10.1)	
Race/ethnicity			<.0001
Non-Hispanic White	105,613 (64.7)	70,340 (69.3)	
Non-Hispanic Black	23,421 (14.4)	12,183 (12.0)	
Non-Hispanic Asian/Pacific Islander	10,912 (6.7)	6,006 (5.9)	
Hispanic	20,677 (12.7)	10,566 (10.4)	
Other/missing	2,510 (1.5)	2,379 (2.3)	
Marital status			<.0001
Single	36,468 (22.4)	19,650 (19.4)	
Married	82,958 (50.9)	54,628 (53.8)	
Divorced/separated	16,004 (9.8)	9,760 (9.6)	
Widowed	18,752 (11.5)	9,334 (9.2)	
Other/unknown	8,951 (5.5)	8,102 (8.0)	
Resident of New York City/Long Island	87,956 (53.9)	50,550 (49.8)	<.0001
Deceased	39,402 (24.2)	16,488 (16.3)	<.0001
Stage at diagnosis			<.0001
Local	73,857 (45.3)	50,119 (49.4)	
Regional	34,173 (21.0)	21,917 (21.6)	
Distant	41,967 (25.7)	22,093 (21.8)	
Unknown	13,136 (8.1)	7,345 (7.2)	
Received chemotherapy‡	50,064 (30.7)	32,488 (32.0)	<.0001
Received hormonal therapy‡	26,125 (16.0)	24,083 (23.7)	<.0001
Received immunotherapy‡	14,051 (8.6)	10,578 (10.4)	<.0001
Received radiation therapy‡	46,319 (28.4)	32,541 (32.1)	<.0001
Received surgery‡	79,225 (48.6)	59,665 (58.8)	<.0001

Note: Percentages may not sum to 100 due to rounding.

*ePath reporting defined as an electronic pathology report submitted within 3 months after diagnosis.

†P value from $\chi 2$ test.

‡Based on consolidated data on treatment received during initial therapy; treatment categories are not mutually exclusive.

Looking at each diagnosis year separately, the differences in case characteristics between cases diagnosed vs those not diagnosed by ePath were similar to those for all 2017–2019 cases combined. However, in 2019 the differences were attenuated but still statistically significant for race/ethnicity (66.3% vs 64.4% non-Hispanic White), while for region of residence the direction of the relationship changed such that in 2019 ePath cases were more likely to be residents of NYC or Long Island than cases not reported by ePath (54.3% vs 50.8%; results not shown). These changes were likely related to the onboarding of additional ePath reporting facilities in NYC in 2019.

Table 2. Number and Percent of Cancer Cases Reported by ePath by Site and Diagnosis Year for 264,607 Malignant First Primary Cancers Diagnosed in New York State Residents Aged 18 Years and Older in 2017–2019 and Reported to the New York State Cancer Registry*

Cancer site	*Diagnosis year*			
	2017	**2018**	**2019**	**2017–2019 combined**
All sites combined	30,488 (34.6)	31,095 (35.5)	39,891 (44.8)	101,474 (38.3)
Bladder	401 (22.2)	474 (26.8)	627 (35.0)	1,502 (28.0)
Breast	6,765 (50.5)	6,884 (50.7)	8,253 (59.1)	21,902 (53.5)
Colon/rectum	2,035 (27.1)	2,046 (27.9)	2,923 (39.6)	7,004 (31.5)
Kidney	654 (21.9)	765 (24.9)	1,196 (37.9)	2,615 (28.4)
Leukemia	1,328 (52.3)	1,178 (48.5)	1,197 (51.0)	3,703 (50.6)
Lung	2,330 (23.0)	2,516 (25.5)	3,706 (37.9)	8,552 (28.7)
Lymphoma	1,783 (40.9)	1,801 (42.1)	2,175 (49.8)	5,759 (44.3)
Melanoma	1,926 (63.0)	1,925 (64.0)	2,169 (69.1)	6,020 (65.4)
Myeloma	617 (38.9)	584 (38.2)	653 (42.0)	1,854 (39.7)
Ovary	278 (22.5)	284 (25.2)	423 (36.5)	985 (28.0)
Pancreas	643 (23.7)	698 (26.1)	1,065 (39.9)	2,406 (29.9)
Prostate	4,556 (34.0)	4,789 (33.8)	5,819 (39.4)	15,164 (35.8)
Stomach	450 (27.3)	443 (28.6)	640 (40.9)	1,533 (32.2)
Thyroid	1,267 (37.8)	1,166 (35.8)	1,624 (47.4)	4,057 (40.5)
Uterus	1,438 (41.5)	1,527 (45.1)	1,796 (52.8)	4,761 (46.4)
Other/miscellaneous	4,017 (26.9)	4,015 (27.7)	5,625 (38.9)	13,657 (31.1)

Data presented as no. (%). *ePath reporting defined as an electronic pathology report submitted within 3 months after diagnosis.

Looking by cancer site, the sites with the highest percentage of cases reported by ePath in 2017–2019 were melanoma (65.4%), breast (53.5%), and leukemia (50.6%), while the sites with the lowest ePath reporting were bladder (28.0%), ovary (28.0%), kidney (28.4%), and lung (28.7%; Table 2). Between 2017–2019, the percent of cases reported by ePath increased for all sites, although the magnitude of the increase varied. The biggest increases were seen for kidney (73.1% increase), pancreatic (68.4% increase), lung (64.8% increase), and ovarian cancer (62.2% increase), although the percent of cases reported by ePath remained relatively low in 2019 for all 4 sites.

In unadjusted analyses, there were some statistically significant differences in the characteristics of cases reported vs not reported by ePath for each cancer site, although these differences and their magnitude varied by site. For example, the distribution of age categories and stage at diagnosis differed significantly by ePath status for each cancer site, but for age the direction of the relationship was fairly consistent across cancer sites while for disease stage the relationship varied by site (results not shown). We used logistic regression to examine these relationships in more detail for common cancers and to adjust for other potential correlates of ePath reporting status.

The odds of ePath reporting were higher for females vs males for all cancers combined (OR, 1.04; 95% CI, 1.02–1.07), but no statistically significant association was observed for colorectal cancer, lung cancer, lymphoma, or myeloma (Table 3), suggesting that the differences observed in the sex distribution of cancer cases by ePath status may be due to the prevalence of ePath reporting for sex-specific cancers. Overall the multivariable-adjusted associations for all cancers combined mirrored the unadjusted associations presented in Table 1 and were statistically significant even after adjustment for other variables. For breast cancer, statistically significant associations with ePath reporting were observed for all variables examined, while for other common cancers some case characteristics were statistically significantly associated with ePath reporting and others were unassociated.

For each cancer site included in Table 3, the odds of ePath reporting were statistically significantly lower for older individuals (ages 70–79 years and/or 80 years and older vs ages 60–69 years), while for colorectal, lung, and prostate cancers the odds of ePath reporting were also statistically significantly higher for younger individuals ages 18–49 years and 50–59 years. When compared to non-Hispanic White individuals, the odds of ePath reporting for breast cancer were statistically significantly lower for non-Hispanic Black women (OR, 0.63; 95% CI, 0.59–0.67), non-Hispanic Asian/Pacific Islanders (OR, 0.61; 95% CI, 0.56–0.66), and Hispanic women (OR, 0.62; 95% CI, 0.58–0.66). Similarly, the odds of ePath reporting for colorectal cancer were lower among Hispanic individuals, and the odds of ePath reporting for lung cancer and lymphoma were lower for both Hispanic individuals and non-Hispanic

Table 3. Multivariable-Adjusted Odds Ratios (OR) and 95% CIs for ePath Reporting within 3 Months after Diagnosis by Case Characteristic, Overall and for Common Cancer Sites, among Malignant First Primary Cancers Diagnosed in New York State Residents Aged 18 Years and Older in 2017–2019 and Reported to the New York State Cancer Registry

Case characteristic	*OR (95% CI)**						
	All cancers	**Breast**	**Colorectal**	**Lung**	**Lymphoma**	**Melanoma**	**Prostate**
Sex							
Male	1.00 (ref)	–	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	–
Female	1.04 (1.02–1.07)	–	1.00 (0.94–1.06)	1.06 (1.00–1.12)	0.95 (0.89–1.02)	1.08 (0.99–1.18)	–
Other/unknown	0.70 (0.41–1.19)	–	0.86 (0.26–2.83)	–	2.88 (0.85–9.74)	–	–
Age category, y							
18–49	1.03 (1.00–1.06)	0.98 (0.93–1.04)	1.17 (1.07–1.29)	1.18 (1.03–1.35)	0.93 (0.83–1.03)	1.11 (0.97–1.27)	1.33 (1.15–1.53)
50–59	1.07 (1.05–1.10)	1.02 (0.97–1.08)	1.12 (1.03–1.21)	1.17 (1.08–1.27)	0.96 (0.86–1.07)	1.07 (0.94–1.22)	1.14 (1.08–1.20)
60–69	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
70–79	0.91 (0.89–0.93)	0.84 (0.79–0.89)	0.94 (0.87–1.03)	0.93 (0.87–0.99)	0.86 (0.77–0.95)	0.96 (0.85–1.10)	0.87 (0.83–0.91)
≥80	0.73 (0.71–0.76)	0.66 (0.61–0.72)	0.78 (0.70–0.86)	0.78 (0.72–0.85)	0.77 (0.68–0.88)	0.84 (0.72–0.97)	0.66 (0.60–0.72)
Race/ethnicity							
Non-Hispanic White	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Non-Hispanic Black	0.85 (0.83–0.87)	0.63 (0.59–0.67)	0.99 (0.91–1.08)	0.94 (0.86–1.02)	0.91 (0.81–1.03)	0.69 (0.42–1.14)	0.99 (0.94–1.05)
Non-Hispanic Asian/ Pacific Isl.	0.85 (0.82–0.88)	0.61 (0.56–0.66)	1.02 (0.91–1.15)	0.76 (0.68–0.86)	0.69 (0.59–0.82)	1.40 (0.79–2.48)	1.33 (1.20–1.48)
Hispanic	0.80 (0.78–0.82)	0.62 (0.58–0.66)	0.82 (0.75–0.90)	0.83 (0.74–0.92)	0.83 (0.75–0.93)	0.82 (0.63–1.07)	1.02 (0.96–1.10)
Other/Missing	1.19 (1.12–1.26)	0.82 (0.68–0.97)	0.93 (0.72–1.20)	1.14 (0.84–1.55)	0.82 (0.62–1.09)	1.82 (1.41–2.36)	1.54 (1.36–1.75)
Marital status							
Single	0.86 (0.84–0.88)	0.82 (0.77–0.86)	0.78 (0.73–0.85)	0.89 (0.83–0.96)	1.00 (0.91–1.10)	1.04 (0.91–1.19)	0.82 (0.77–0.87)
Married	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Divorced/separated	0.94 (0.92–0.97)	0.92 (0.86–0.98)	0.97 (0.88–1.07)	0.96 (0.88–1.04)	1.05 (0.92–1.20)	0.88 (0.74–1.05)	0.87 (0.81–0.95)
Widowed	0.92 (0.89–0.95)	0.91 (0.85–0.97)	0.90 (0.82–1.00)	0.90 (0.83–0.97)	0.99 (0.86–1.13)	1.02 (0.84–1.23)	0.85 (0.76–0.96)
Other/unknown	1.39 (1.35–1.45)	0.97 (0.88–1.08)	0.97 (0.85–1.11)	0.85 (0.72–0.99)	1.33 (1.12–1.57)	1.86 (1.62–2.13)	1.31 (1.21–1.42)
Region of residence							
New York City/ Long Island	0.89 (0.88–0.91)	0.91 (0.87–0.95)	1.06 (0.99–1.12)	0.74 (0.69–0.78)	0.65 (0.60–0.70)	0.75 (0.69–0.82)	1.30 (1.24–1.36)
Rest of state	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Missing/unknown	0.67 (0.46–0.97)	0.61 (0.29–1.31)	–	0.76 (0.07–8.94)	0.15 (0.02–1.22)	2.27 (0.27–18.8)	1.38 (0.63–3.03)

Table 3, cont. Multivariable-Adjusted Odds Ratios (OR) and 95% CIs for ePath Reporting within 3 Months after Diagnosis by Case Characteristic, Overall and for Common Cancer Sites, among Malignant First Primary Cancers Diagnosed in New York State Residents Aged 18 Years and Older in 2017–2019 and Reported to the New York State Cancer Registry

Case characteristic	*OR (95% CI)**						
	All cancers	**Breast**	**Colorectal**	**Lung**	**Lymphoma**	**Melanoma**	**Prostate**
Stage							
Local	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Regional	1.03 (1.01–1.05)	0.84 (0.79–0.88)	1.00 (0.93–1.08)	1.11 (1.02–1.20)	1.20 (1.06–1.35)	0.95 (0.80–1.13)	1.12 (1.05–1.20)
Distant	1.03 (1.00–1.06)	0.68 (0.61–0.76)	0.76 (0.68–0.84)	1.30 (1.20–1.41)	1.30 (1.19–1.43)	0.41 (0.31–0.55)	0.64 (0.57–0.71)
Unknown	1.03 (0.99–1.06)	1.07 (0.92–1.25)	0.77 (0.67–0.88)	0.62 (0.51–0.76)	1.20 (1.05–1.38)	0.89 (0.76–1.05)	1.36 (1.26–1.46)
Received chemotherapy†							
No/unknown	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Yes	1.08 (1.06–1.11)	0.78 (0.74–0.82)	1.26 (1.17–1.36)	1.05 (0.99–1.12)	0.95 (0.87–1.05)	1.65 (1.14–2.40)	1.07 (0.88–1.29)
Received hormonal therapy†							
No/unknown	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Yes	1.15 (1.12–1.18)	1.32 (1.26–1.38)	1.70 (1.10–2.61)	1.15 (0.85–1.56)	0.92 (0.83–1.01)	0.21 (0.02–1.96)	0.97 (0.92–1.03)
Received immunotherapy†							
No/unknown	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Yes	1.12 (1.09–1.16)	0.91 (0.85–0.98)	1.55 (1.35–1.77)	1.33 (1.23–1.43)	1.36 (1.24–1.49)	1.09 (0.88–1.35)	1.47 (1.06–2.04)
Received radiation therapy†							
No/unknown	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Yes	1.11 (1.09–1.14)	1.13 (1.08–1.19)	0.90 (0.82–0.98)	1.21 (1.14–1.29)	1.13 (1.02–1.26)	0.99 (0.72–1.35)	0.82 (0.78–0.86)
Received surgery†							
No/unknown	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Yes	1.41 (1.38–1.44)	1.46 (1.34–1.58)	0.98 (0.90–1.06)	1.86 (1.72–2.00)	1.09 (0.99–1.20)	0.97 (0.82–1.15)	1.17 (1.11–1.23)
Diagnosis year							
2017	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
2018	1.04 (1.02–1.06)	1.01 (0.96–1.06)	1.04 (0.97–1.12)	1.12 (1.05–1.20)	1.05 (0.96–1.15)	1.04 (0.93–1.15)	1.01 (0.96–1.06)
2019	1.58 (1.55–1.61)	1.53 (1.46–1.61)	1.77 (1.65–1.90)	2.00 (1.88–2.13)	1.45 (1.33–1.58)	1.26 (1.13–1.40)	1.25 (1.19–1.32)

*Adjusted for sex (if applicable), age, race/ethnicity, marital status, region of residence, cancer site (for analysis of all cancers), stage at diagnosis, diagnosis year, and treatment received during initial therapy based on consolidated data.

†Based on consolidated data on treatment received during initial therapy; treatment categories are not mutually exclusive.

Asian/Pacific Islanders. In contrast, the odds of ePath reporting for prostate cancer were statistically significantly higher for non-Hispanic Asian/Pacific Islanders (OR, 1.33; 95% CI, 1.20–1.48) when compared to non-Hispanic White individuals. The odds of ePath reporting were lower for single vs married individuals for breast, colorectal, lung, and prostate cancers and additionally were lower for divorced/separated individuals for breast and prostate cancers and for widowed individuals for breast, lung, and prostate cancers. The odds of ePath reporting were lower for patients residing in NYC/Long Island vs upstate for all cancers examined except colorectal cancer, which was unassociated with region of residence, and prostate cancer, where the odds of ePath reporting were higher for NYC/

Long Island residents (OR, 1.30; 95% CI, 1.24–1.36). The opposite direction of the association for prostate cancer is likely related to differences in the facilities where prostate cancer patients tend to be diagnosed and treated, when compared with other cancers.

Similarly, there were statistically significant differences in disease stage and treatment by ePath reporting status for all cancer sites, although the associations were inconsistent across cancers. The odds of ePath reporting were lower for regional and/or distant stage disease for breast cancer, colorectal cancer, and melanoma, while for lung cancer and lymphoma the odds of ePath reporting were higher for regional and distant vs local stage cancers. For prostate cancer, the odds of ePath reporting were higher for regional cancers (OR, 1.12; 95% CI, 1.05–1.20) but lower for distant stage cancers (OR, 0.64; 95% CI, 0.57–0.71), when compared to localized cancers. The odds of ePath reporting tended to be higher among cancers treated vs not treated with particular modalities during initial therapy. Receipt of chemotherapy was associated with lower odds of ePath reporting for breast cancer (OR, 0.78; 95% CI, 0.74–0.82), higher odds of ePath reporting for colorectal cancer (OR, 1.26; 95% CI, 1.17–1.36) and melanoma (OR, 1.65; 95% CI, 1.14–2.40), and was unassociated with ePath reporting for the other cancers examined. Receipt of hormonal therapy was associated with increased odds of ePath reporting for breast and colorectal cancers only, while receipt of immunotherapy was associated with increased odds of ePath reporting for colorectal cancer, lung cancer, lymphoma, and prostate cancer and decreased odds of ePath reporting for breast cancer. Receipt of radiation therapy was inconsistently associated with ePath reporting, with increased odds of ePath reporting for breast cancer, lung cancer, and lymphoma, decreased odds for colorectal and prostate cancers, and no association for melanoma. Receipt of surgery was associated with increased odds of ePath reporting for breast (OR, 1.46; 95% CI, 1.34–1.58), lung (OR, 1.86; 95% CI, 1.72–2.00), and prostate cancers only (OR, 1.17; 95% CI, 1.11–1.23). For all cancer sites, the odds of ePath reporting were statistically significantly higher for diagnosis year 2019 vs 2017, reflecting increases over time in the number of facilities with ePath reporting capabilities.

Discussion

Our results show that ePath reporting to the NYSCR has increased over time but remains incomplete, introducing the potential for bias in patient contact studies that rely on ePath reporting for rapid case ascertainment. For the most recent diagnosis years with complete data, we observed statistically significant differences between cases reported vs not reported by ePath for all variables examined. These differences were present for all cancers combined and for common cancer sites, and they persisted after adjusting for other case characteristics. Although the percentage of cases reported by ePath was higher for some cancer sites than others, we observed statistically significant differences by ePath status regardless of the proportion of cases reported by ePath. Some results suggested that cancers with a more aggressive diagnostic workup and treatment protocol, as would be expected in younger individuals, were more likely to be reported by ePath. These results indicate that studies that use ePath reports for rapid case ascertainment, without additional case identification approaches to capture cases not reported by ePath, will no longer be representative of the underlying population.

Rapid ascertainment of newly diagnosed cancer cases is often of interest in patient contact research studies to minimize the time between a patient's diagnosis and study enrollment.[4-6] Potential benefits include assessment of prediagnostic exposures closer to the time of diagnosis, which may decrease recall bias, and prospective assessment of clinical and treatment-related factors including treatment decision-making, barriers to treatment, and quality of life. In addition, in studies of highly fatal cancers and transient populations, rapid case ascertainment may maximize participation and reduce the number of patients who are unable to be reached.[7] Potential drawbacks of rapid case ascertainment include that patients may not be emotionally ready to participate in a research study soon after receiving a cancer diagnosis or may be too ill to participate due to symptoms of their cancer or adverse effects of active treatment. In addition, rapid case ascertainment adds to the work required by a cancer registry, since these studies necessitate review and eligibility assessment of ePath reports earlier than would be the case during the normal registry review process, adding to the cost and logistics involved. These early reports also contain fewer data items and do not go through the usual registry quality control processes at the time of initial report, likely resulting in incomplete data and some quality issues in samples identified by rapid case ascertainment.

Despite these challenges, in many studies, the advantages of rapid case ascertainment in a cancer registry may outweigh the limitations, in particular when enrollment in the study and initial data collection are time sensitive. Other possible sources of cancer cases, such as large cancer centers or insurance companies, often are not representative of the underlying population, since certain groups of patients including those with limited financial resources may be underrepresented.[1] As a result, these studies are vulnerable to selection bias and study findings may not be generalizable to subpopulations not represented in the study sample. Studies using state and regional cancer registries for case identification have the potential to be population-based, since almost all cancer cases occurring in the population are captured due to mandated reporting and thorough follow-up. However, the lag time between diagnosis and availability of complete case information may be prohibitive in some studies.[8] As observed in this analysis, relying on incomplete ePath reporting to central cancer registries can result in similar issues with selection bias and generalizability that occur with other sources of cancer cases. Although these issues do not necessarily affect the internal validity of a study, approaches to maximize the representativeness of the sample and the external validity of the study are needed.

In New York State, the absolute percentage of cases reported by ePath within 3 months after diagnosis increased an average of 5.8% per year between 2014–2019. However,

despite this increase and the higher proportion of cases reported by ePath for certain cancer sites, we observed statistically significant differences between the ePath cases and cases not reported by ePath. The NYSCR is continuing to work with additional facilities to enable them to submit ePath reports; however, the onboarding process is slow and requires commitment and dedicated resources by the facility for implementation and continued support of ePath reporting. This long-term commitment can be cost prohibitive, in particular for smaller or under-resourced reporting facilities. Approaches to streamline the ePath reporting process and to minimize the cost of ePath reporting will benefit both cancer registries and the entire cancer research community.

Strengths of this analysis include the availability of complete NYSCR data for diagnosis years 2014–2019, including information on the date and format of each case report. The detailed data available on patient demographics, clinical characteristics, and treatment received during initial therapy allowed us to compare multiple characteristics of the cases reported by ePath and those not reported by ePath, and the use of multivariable-adjusted logistic regression allowed for control for other possible correlates of ePath reporting.

Our analysis was limited to data from a single state cancer registry, which may not be representative of ePath reporting and differences in patient characteristics by ePath status in other states. However, New York is a diverse state with a mix of urban and rural areas and both large cancer centers and smaller reporting facilities, suggesting that the results may be relevant to other states with similar patterns of ePath reporting. In addition, the variables examined were limited to those routinely collected by cancer registries in the United States and may not include other variables of interest to researchers for assessing the representativeness of cases reported by ePath.

In summary, our results indicate that relying solely on incomplete ePath reporting for identification of cancer cases for patient contact research studies will result in a sample that differs from the underlying population, potentially resulting in selection bias and limiting the generalizability of the study results. Work is needed to increase the proportion of reporting facilities with ePath reporting capability, as well as the accessibility of ePath reporting software. In the meantime, approaches to include cases not reported by ePath in rapid case ascertainment studies, such as by working directly with reporting facilities without ePath to identify additional cases or extending the time window for identification and enrollment of cases, will help to improve the representativeness of the study sample.

References

1. Tucker TC, Durbin EB, McDowell JK, Huang B. Unlocking the potential of population-based cancer registries. *Cancer.* 2019;125(21):3729-3737.
2. Hill SM, Li J, Pawlish K, Paddock LE, Stroup AM. Unintended consequences of expanding electronic pathology reporting: the inverse relationship between data completeness and data quality. *J Registry Manag.* 2020;47(3):122-126.
3. North American Association of Central Cancer Registries, Inc. *2018 Implementation Guidelines and Recommendations.* Version 1.2. Revised November 2018. Accessed August 24, 2021. **https://www.naaccr.org/wp-content/uploads/2020/03/2018-Implementation-Guidelines20181101a.pdf**
4. Aldrich TE, Vann D, Moorman PG, Newman B. Rapid reporting of cancer incidence in a population-based study of breast cancer: one constructive use of a central cancer registry. *Breast Cancer Res Treat.* 1995;35(1):61-64.
5. Pearson ML, Ganz PA, McGuigan K, Malin JR, Adams J, Kahn KL. The case identification challenge in measuring quality of cancer care. *J Clin Oncol.* 2002;20(21):4353-4360.
6. Ryerson AB, Eheman C, Styles T, Rycroft R, Snyder C. Connecting the dots: linking the National Program of Cancer Registries and the needs of survivors and clinicians. *Am J Prev Med.* 2015;49(6 suppl 5):S528-S535.
7. Deapen D. Thirty-three years of rapid case ascertainment: lessons learned [PowerPoint slides]. Presented at NAACCR Annual Meeting; June 8, 2005; Cambridge, MA. Accessed August 20, 2021. **https://www.naaccr.org/wp-content/uploads/2016/11/Thirty-three-Years-of-Rapid-Case-Ascertainment-Lessons-Learned.pdf**
8. About the New York State Cancer Registry. New York State Department of Health website. Accessed August 20, 2021. **https://www.health.ny.gov/statistics/cancer/registry/about.htm**

Original Article

Predictors of Survival among Male and Female Patients with Malignant Pleural Mesothelioma: A Random Survival Forest Analysis of Data from the 2000–2017 Surveillance, Epidemiology, and End Results Program

Bian Liu, PhD[a]; Li Niu, PhD[a]; Francis Boscoe, PhD[b, c]; Furrina F. Lee, PhD[c]

Abstract: Background: Malignant pleural mesothelioma (MPM) is a rare and aggressive malignancy with a dismal prognosis. We aimed to identify predictors of survival among male and female MPM patients in the United States. Methods: We identified MPM cases reported by 18 cancer registries in the Surveillance, Epidemiology, and End Results Program (2000–2017). We applied a random survival forest (RSF) algorithm to identify and rank the importance of 10 variables at patient, cancer, and area level in predicting all-cause survival overall and by female and male subgroups. Results: Approximately 91.4% (n = 11,160) of the MPM patients had died, with better survival among females than males (11.7% vs 7.8%). The median follow-up time was 7 months (interquartile range, 2–17 months). A majority of the patients were male (78.6%), non-Hispanic White (81.8%), and residing in metropolitan counties with a population greater than 1 million (63.7%). The top 3 factors for predicting overall MPM survival were age, histological type, and cancer-directed surgery status. Except for age, the relative ranking of covariates varied by the 3 sample groups. Stage ranked fifth in predicting female survival, while it was replaced by metastasis status for male and overall patients. Race/ethnicity was not a good predictor for survival among MPM patients overall or the male subgroup, but ranked sixth for predicting survival among females. Median household income was not a good predictor for survival among females. Conclusion: We demonstrated that RSF successfully identified predictors of MPM survival. RSF is a viable complement to the commonly used Cox proportional hazard model and a viable alternative, particularly when the proportional hazard assumption is unmet. RSF also identified differences between the sexes, which may help explain the sex differences in MPM survival rates.

Key words: *machine learning; mesothelioma; Surveillance, Epidemiology, and End Results (SEER) Program; survival; variable importance*

Introduction

Malignant pleural mesothelioma (MPM) is a rare and aggressive malignancy that develops from the thin layer of tissues surrounding the lungs and chest walls.[1,2] MPM is the dominant type of mesothelioma (accounting for approximately 90% of cases), with known carcinogenesis of asbestos exposure linked to occupation (70%–80% of all cases of mesothelioma) or environmental exposure.[2-4] The 5-year survival rate for patients with MPM is less than 10% because cases are often diagnosed at a late stage and treatment options are generally not curative.[5]

Previous population-based observational studies often used traditional survival analysis, such as the Cox proportional hazard (CPH) regression model, to identify and evaluate factors associated with survival.[6-12] Predictors of survival varied between studies and the selection of predictors was often subjective. Because factors contributing to poor survival—such as stage, treatment types, age, sex, race/ethnicity, and socioeconomic status—are interrelated, multicollinearity could be a concern. Moreover, the proportional hazard assumption for the commonly used CPH regression model for survival analysis is often unmet. In contrast, machine-learning methods, such as the random survival forest (RSF), can address the aforementioned gaps. Moreover, machine-learning methods are better equipped than traditional survival regression models to handle complex nonlinear relationships among large multidimensional data with both categorical and continuous variables.

There is a growing interest in using machine-learning models to analyze oncological survival data,[13-17] though their application in mesothelioma is limited. The present study adds to the existing literature by applying RSF to explore the complex relationship between MPM survival and its correlates at the levels of patient, cancer, and

[a] Icahn School of Medicine at Mount Sinai, New York, New York. [b] Pumphandle, LLC, Portland, Maine. [c] Bureau of Cancer Epidemiology, Division of Chronic Disease Prevention, New York State Department of Health, Menands, New York.

Address correspondence to Bian Liu, PhD, Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1077, New York, NY 10029. Telephone: (212) 659-1451. Fax: (212) 423-2998. Email: bian.liu@mountsinai.org.

This work was supported in part by a grant from the National Cancer Institute (1R21CA235153).

All the data used for these analyses had been deidentified by the Surveillance Epidemiology and End Results program (SEER), supported by the National Cancer Institute. The deidentified data were made available to researchers with a registered SEER*Stat account. The use and reporting of the data comply with the Data-Use Agreement for the 1975–2017 SEER Research Data File. Given that our study uses these data that cannot be linked to specific individuals either directly or indirectly, and were not collected specifically for the currently proposed research project through any interaction with the patients, this project is considered not human subjects research.

geography. We used the most recent data from 18 registries in the Surveillance, Epidemiology, and End Results (SEER) Program, representing 27.8% of the US population.[18] Supported by the National Cancer Institute, SEER provides quality data for researchers to investigate cancer control at the population level. Our results may provide new insight into the relative importance of factors in predicting the survival of MPM patients overall and by sex.

Methods

Data Source

Using SEER*Stat software, we extracted the study population from the incidence data reported by 18 registries (2000–2017) within SEER.[19] This database contained the most recent data at the time of this writing.[20]

Of the 15,047 cases with mesothelioma as the cancer site (site recode ICD-O-3/WHO 2008 = "Mesothelioma", primary site: C384 and histology: 9050–9053), we excluded cases with unknown diagnostic confirmation (n = 415), unknown survival time (n = 56), and where mesothelioma was not the primary cancer site (n = 32), as well as mesothelioma types other than MPM (n = 2,291). We further excluded 43 cases due to missing information on at least 1 of the covariates (see details below). The final sample included 12,210 MPM patients.

Outcome and Covariate Measures

Overall survival (vs death from all causes) was the outcome of interest. Survival status and follow-up time in months were taken directly from the downloaded SEER research database, which was based on the November 2019 submission. We included the following 10 variables based on the availability of the data: 5-year age group (<55, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, and ≥85 years), sex (male and female), race/ethnicity (non-Hispanic White, non-Hispanic Black, Hispanic, and other), county-level median household income (<$45K, $45K–$54K, $55K–$64K, $65K–$74K, >$75K), county-level rural-urban continuum code (nonmetropolitan areas, adjacent to a metropolitan area, metropolitan with a population <250K, metropolitan with a population between 250K and 1 million, and metropolitan with a population >1 million), histology (9050/3: mesothelioma, 9051/3: fibrous, 9052/3: epithelioid, and 9053/3: biphasic), summary stage (local, regional, distant, and unknown), metastases (yes/no), cancer-directed surgery status (not recommended, recommended but not performed, and performed), and year of diagnosis. Cancer metastases was defined as distant metastases in bone, brain, liver, lung, distant lymph nodes, or other.

Statistical Analysis

To identify and rank the importance of covariates in predicting survival, we applied the RSF algorithm developed by Ishwaran and Kogalur.[21] RSF expands the existing random forest algorithm to accommodate time-to-event data. RSF is a common machine-learning algorithm based on classification and regression tree methods, which classify data or populations according to covariates through a recursive partitioning and a validation process.[22] RSF is an ensemble of classification or regression trees (for categorical and continuous outcomes), where each tree is based on a random subset of predictors and outcome at each node, each tree "votes" for the classification of the outcome (ie, survival), and the final classification is based on the most votes over all the trees in the forests. In the process, the method also produces a rank of the importance of the variables that can be used for dimension reduction. The nonparametric design (without strong assumptions of distributions of variables as in traditional regression models) enables RSF to handle complex nonlinear relationships among large multidimensional data with both categorical and continuous variables and address concerns of collinearity among multiple covariates.

We implemented RSF using the randomForestSRC R package.[23] We grew 1,000 survival classification trees and used the default log-rank splitting rule to maximize survival differences between child nodes. The number of random split points was 10 and the number of variables randomly selected for splitting a node was 4 and 3 for the overall and sex-specific analyses, respectively. We used the default Gini index splitting rules to obtain the variable importance (VIMP) based on the Breiman–Cutler permutation variable importance, where a positive or negative VIMP value indicate that a variable improves or decreases the prediction. The error rate was based on Harrell's concordance index, which uses the cumulative hazard estimate as the value for comparison. The error rate (1-C) ranges from 0 (perfect prediction) to 1, and a value of 0.5 corresponds to a prediction no better than random guessing.

We also compared predictions between the RSF and CPH models using the default Kaplan–Meier model as the benchmark. The evaluation was conducted by calculating the prediction errors from a bootstrap (n = 500) cross-validation. The integrated Brier score (IBS), which is a cumulative prediction error, was calculated. Brier score measures the accuracy of a predicted survival function at a given time, *t*, which ranges from 0 to 1 with zero being the best possible value. IBS represents an overall model performance at all available times.

To help interpret the findings from RSF, we applied the CPH model to obtain the hazard ratios (HRs) and their 95% CIs using only the top 5 predictors from the RSF. The purpose of this step is to quantify the effects of individual predictors on survival to provide a familiar and clinically understandable output. We also checked the proportional hazard assumption. To visualize how a predictor might vary over time, we applied Aalen's additive regression model to the same set of the top 5 predictors and plotted the cumulative regression effects, which are regression β coefficients (ie, slopes) and their 95% CIs as a function of time.[24] These 2 analysis steps were implemented using the survival R package. We used SAS (version 9.4) to conduct the initial data preparation of the exported datasets from SEER*Stat, and used R (version 3.5.0) with RStudio (version 1.1.453) for all statistical analyses.

Table 1. Study Population Characteristics				
Variables		*Overall, n (%)*	*Female, n (%)*	*Male, n (%)*
Survival status	Alive	1,050 (8.6)	306 (11.7)	744 (7.8)
	Dead	11,160 (91.4)	2,306 (88.3)	8,854 (92.2)
Follow-up time (months)	Mean (SD)	13.6 (20.0)	16.67 (24.6)	12.8 (18.5)
	Median (interquartile range)	7 (2-17)	8 (3-20)	7 (2–16)
	Minimum–maximum	0–210	0–205	0–210
Age (y)	<55	779 (6.4)	295 (11.3)	484 (5.0)
	55–59	709 (5.8)	188 (7.2)	521 (5.4)
	60–64	1,117 (9.1)	255 (9.8)	862 (9.0)
	65–69	1,562 (12.8)	301 (11.5)	1,261 (13.1)
	70–74	2,009 (16.4)	368 (14.1)	1,641 (17.1)
	75–79	2,329 (19.1)	426 (16.3)	1,903 (19.8)
	80–84	2,074 (17.0)	432 (16.5)	1,642 (17.1)
	≥85	1,631 (13.4)	347 (13.3)	1,284 (13.4)
Sex	Female	2,612 (21.3)	2,612 (100%)	
	Male	9,598 (78.6)		9,598 (100%)
Race/ethnicity	Non-Hispanic White	9,993 (81.8)	2,092 (80.1)	7,901 (82.3)
	Non-Hispanic Black	602 (4.9)	143 (5.5)	459 (4.8)
	Hispanic	1,189 (9.7)	267 (10.2)	922 (9.6)
	Other	426 (3.5)	110 (4.2)	316 (3.3)
County-level median household income (USD)	<$45K	816 (6.7)	164 (6.3)	652 (6.8)
	$45K–$54K	1,554 (12.7)	309 (11.8)	1,245 (13.0)
	$55K–$64K	3,238 (26.5)	723 (27.7)	2,515 (26.2)
	$65K–$74K	2,560 (21.0)	512 (19.6)	2,048 (21.3)
	>$75K	4,042 (33.1)	904 (34.6)	3,138 (32.7)
County-level rural-urban Continuum	Nonmetropolitan areas	439 (3.6)	90 (3.4)	349 (3.6)
	Adjacent to a metropolitan area	758 (6.2)	155 (5.9)	603 (6.3)
	Metropolitan with a population <250K	913 (7.5)	168 (6.4)	745 (7.8)
	Metropolitan with a population between 250K and 1 million	2326 (19.0)	434 (16.6)	1,892 (19.7)
	Metropolitan with a population >1 million	7774 (63.7)	1,765 (67.6)	6,009 (62.6)
Cancer stage	Unknown	3148 (25.8)	653 (25.0)	2,495 (26.0)
	Local	934 (7.6)	204 (7.8)	730 (7.6)
	Regional	1534 (12.6)	319 (12.2)	1,215 (12.7)
	Distant	6594 (54.0)	1,436 (55.0)	5,158 (53.7)
Metastases	No	11458 (93.8)	2,447 (93.7)	9,011 (93.9)
	Yes	752 (6.2)	165 (6.3)	587 (6.1)
Histology	9050/3: Mesothelioma	5907 (48.4)	1,288 (49.3)	4,619 (48.1)
	9051/3: Fibrous	1294 (10.6)	170 (6.5)	1,124 (11.7)
	9052/3: Epithelioid	4162 (34.1)	997 (38.2)	3,165 (33.0)
	9053/3: Biphasic	847 (6.9)	157 (6.0)	690 (7.2)

Table 1, cont. Study Population Characteristics				
Variables		*Overall, n (%)*	*Female, n (%)*	*Male, n (%)*
Reason of cancer directed surgery	Surgery not recommended	8492 (69.5)	1,818 (69.6)	6,674 (69.5)
	Surgery recommended but not performed	833 (6.8)	176 (6.7)	657 (6.8)
	Surgery performed	2885 (23.6)	618 (23.7)	2,267 (23.6)
Year of cancer diagnosis	2000–2001	1371 (11.2)	277 (10.6)	1,094 (11.4)
	2002–2003	1324 (10.8)	271 (10.4)	1,053 (11.0)
	2004–2005	1371 (11.2)	255 (9.8)	1,116 (11.6)
	2006–2007	1319 (10.8)	301 (11.5)	1,018 (10.6)
	2008–2009	1439 (11.8)	302 (11.6)	1,137 (11.8)
	2010–2011	1385 (11.3)	277 (10.6)	1,108 (11.5)
	2012–2013	1344 (11.0)	301 (11.5)	1,043 (10.9)
	2014–2015	1355 (11.1)	315 (12.1)	1,040 (10.8)
	2016–2017	1302 (10.7)	313 (12.0)	989 (10.3)

Results

Study Sample Characteristics

Approximately 91.4% (n = 11,160; Table 1) of the MPM patients had died by the study cutoff time. The mean follow-up time was 13.6 months (SD, 20.0 months) and the median was 7 months (interquartile range, 2–17 months). Females had a longer median follow-up time (8 vs 7 months) and better survival than males (11.7% vs 7.8%).

At cancer diagnosis, 80% of the patients were aged 65 years and above. A majority of the patients were male (78.6%), non-Hispanic White (81.8%), and residing in metropolitan counties with a population greater than 1 million (63.7%). About a third of the patients resided in counties where the median household income was greater than $75,000. At diagnosis, 54.0% of the patients had distant cancer and 6.2% had metastases. The epithelioid, fibrous, and biphasic morphological types represented 34.1%, 10.6%, and 6.9% of the patients, respectively, while the remainder were histological code 9050/3 (mesothelioma, malignant), which includes not otherwise specified (NOS) malignant mesothelioma. Surgery was not recommended for close to 70% of the patients, and nearly a quarter of the patients had surgery. The yearly case count within the study period remained relatively stable, ranging from a low of 1,302 in 2016–2017 to a high of 1,439 in 2008–2009. The relative proportions seen in the overall data were largely mirrored among males and females (Table 1). There were also some notable sex differences. For example, compared to males, female patients tended to be younger (eg, 11.3% vs 5% for age < 55 years) and with a more recent diagnosis (~12% vs ~10% for 2014–2017). A higher proportion of females also lived in metropolitan areas with a population over 1 million (67.6% vs 62.6%) and had epithelioid histology (38.2% vs 33.0%).

Variable Importance Ranking

Figure 1 shows the prediction error and VIMP for the overall population and by male and female subgroups. In the overall data, all 10 variables except for race/ethnicity had positive VIMP, indicating they improved the survival prediction. The 9 predictors (from high to low importance) were age, histology, surgery status, diagnosis year, metastasis, summary stage, rural-urban continuum, median household income, and sex. The VIMP ranking in the male subgroup was similar to the results seen in the overall data. Among females, the variable ranking was age, surgery status, histology, summary stage, diagnosis year, race/ethnicity, rural-urban continuum, metastasis, and median household income.

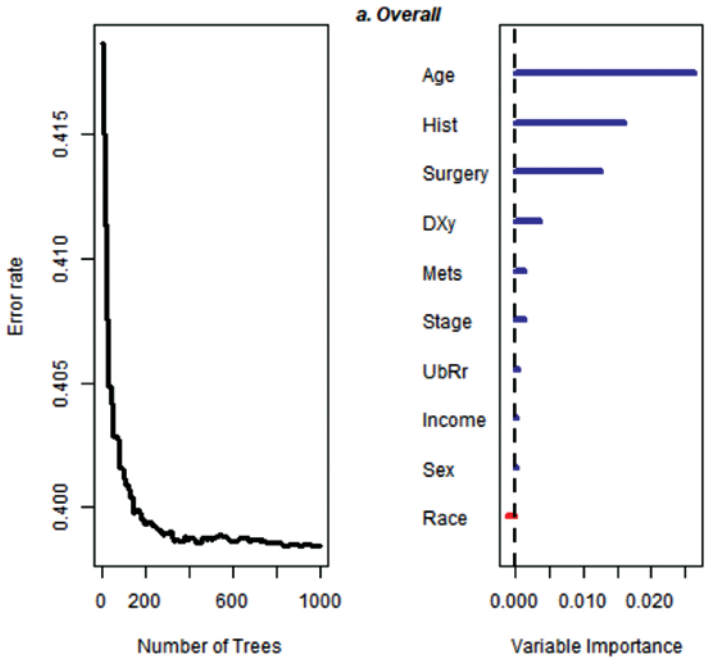
The error rate, based on Harrell's concordance-index (a measure of the model discriminative performance) was 39.85%, 37.58%, and 40.25% for all, female, and male patients, respectively. Using IBS—a measure of both the discrimination and calibration of the model[25]—prediction error from the bootstrap cross-validation showed similar performances across the 3 models. Among the overall sample, the IBS for the Kaplan–Meier, CPH, and RSF models was 0.053, 0.049, and 0.052, respectively. Among females, these scores were 0.074, 0.070, and 0.070; and among males, 0.052, 0.049, and 0.050.

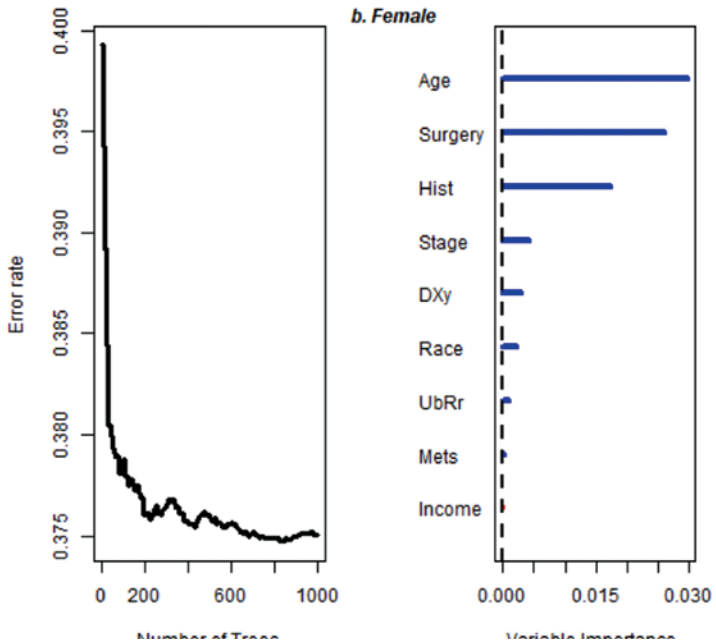
Hazard Ratio Estimates from the Top 5 Predictors

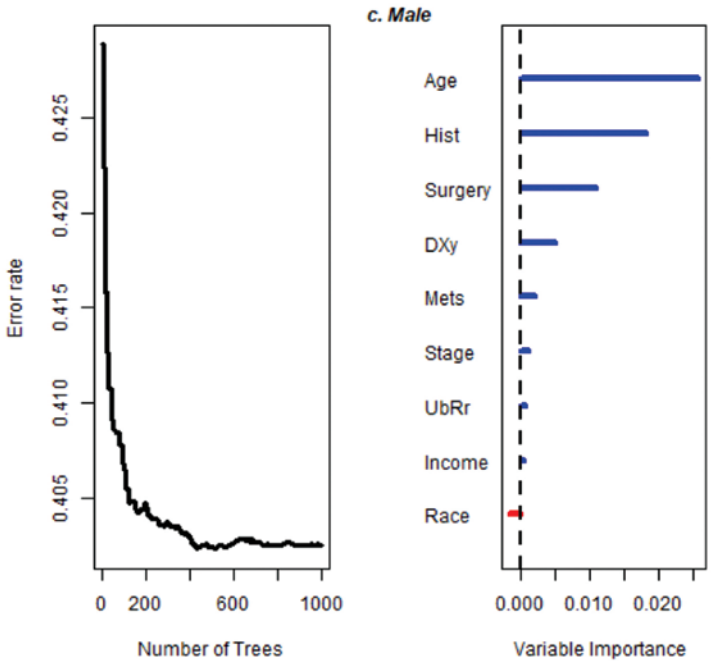
Table 2 presents the association between the top 5 predictors and survival based on the CPH model for all patients, female patients, and male patients. Factors such as age, histology, and diagnosis year showed generally similar relationships with survival in the overall sample and between males and females. Increasing age and having fibrous or biphasic histology subtypes were associated with increased risk of mortality, while having epithelioid histology and being diagnosed more recently were associated with decreased mortality risk. Cancer metastasis increased the risk of dying among all patients (HR, 1.69; 95% CI, 1.55–1.84) and male patients (HR, 1.71; 95% CI, 1.55–1.88). Among females, having a distant cancer stage was associated with worse survival (HR, 1.32; 95% CI, 1.05–1.65) than having an unknown stage. Compared to

Figure 1. Ranking of the Relative Importance of Covariates in Predicting Survival among (a) Overall, (b) Female, and (c) Male Patients

The left side of each panel shows the cumulative error rates as a function of the number of trees, where the out-of-bag error converges/stabilizes quickly as the number of trees in the forest increases. The right panel shows the variable importance (VIMP), where a positive or negative VIMP value indicates that a variable improves or decreases the prediction, after adjusting for all other variables in the model.







the situation where surgery was not recommended, having a recommendation of cancer-directed surgery was statistically significantly associated with a decreased mortality risk among males (and overall patients), regardless of whether surgery was performed or not. However, the association was only significant among females if the surgery was performed (HR, 1.69; 95% CI, 1.55–1.84) and not significant if the surgery was recommended but not performed (HR, 0.87; 95% CI, 0.74–1.03).

The test of the proportional hazard assumption of the CPH model indicated that the assumption was not met for the global model ($P < .0001$) and for individual variables such as age, histology, and surgery status. This was consistent with the time-varying effects of these covariates in predicting survival, as shown in Figure 2. For example, the strong positive slope (y-axis) in age indicated a long-term effect on survival, particularly so for those aged ≥60 years versus 55–59 years. The initial effect of having cancer-directed surgery began to level off close to 30 months among male patients, while it continued to have an effect on survival among female patients.

Discussion

Identifying factors that are strongly associated with mesothelioma mortality is an important area of research given the poor prognosis of mesothelioma. As an alternative to the traditional variable selection to predict survival, we applied a machine-learning algorithm—random survival forest—to 12,210 patients, stratified by sex and in total.

We found that the top 3 most important factors for predicting MPM overall survival were age, histological type, and cancer-directed surgery status. These 3 variables have also been found to be associated with survival in the literature using SEER and other databases.[7-12] For example, epithelioid histology was found to be statistically significantly associated with better survival while biphasic and sarcomatoid histology were not;[11,26] and cancer directed

Table 2. Hazard Ratio (HR) Estimates of the Top 5 Predictors Selected from the Random Survival Forest (RSF) Using Cox Proportional Hazard Models

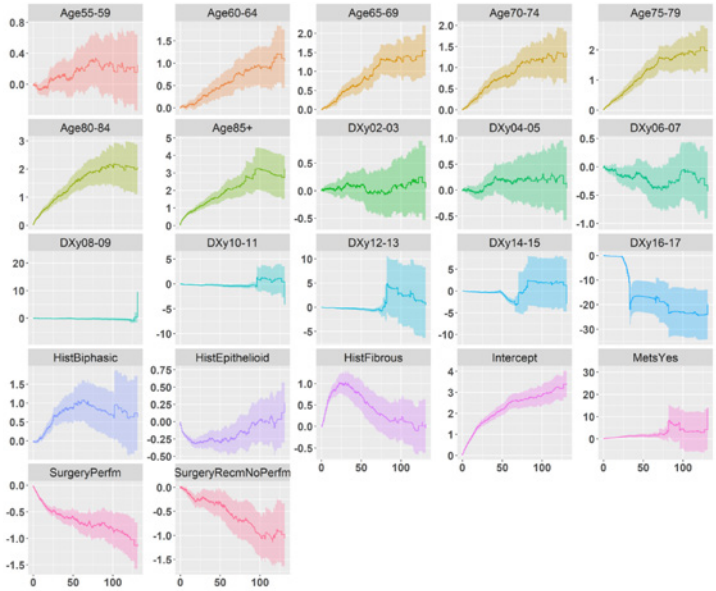
Variables		*Overall HR (95% CI)*	*Female HR (95% CI)*	*Male HR (95% CI)*
Age (y)	<55	Reference	Reference	Reference
	55–59	1.04 (0.93–1.16)	1.13 (0.92–1.39)	0.99 (0.87–1.13)
	60–64	1.25 (1.13–1.38)	1.22 (1.02–1.47)	1.22 (1.08–1.37)
	65–69	1.39 (1.26–1.52)	1.40 (1.17–1.67)	1.32 (1.18–1.48)
	70–74	1.43 (1.30–1.56)	1.45 (1.22–1.72)	1.36 (1.22–1.52)
	75–79	1.61 (1.47–1.76)	1.66 (1.41–1.96)	1.53 (1.37–1.71)
	80–84	1.90 (1.74–2.09)	1.79 (1.52–2.11)	1.87 (1.68–2.09)
	≥85	2.39 (2.18–2.63)	2.34 (1.97–2.79)	2.33 (2.07–2.61)
Cancer stage	Unknown		Reference	
	Local		0.96 (0.73-1.25)	
	Regional		1.22 (0.95-1.57)	
	Distant		1.32 (1.05-1.65)	
Metastases	No	Reference		Reference
	Yes	1.69 (1.55-1.84)		1.71 (1.55-1.88)
Histology	9050/3: Mesothelioma	Reference	Reference	Reference
	9051/3: Fibrous	1.65 (1.55–1.76)	1.76 (1.49–2.08)	1.62 (1.51–1.74)
	9052/3: Epithelioid	0.79 (0.75–0.82)	0.79 (0.72–0.87)	0.79 (0.75–0.83)
	9053/3: Biphasic	1.22 (1.13–1.31)	1.33 (1.11–1.60)	1.18 (1.08–1.28)
Reason of cancer directed surgery	Surgery not recommended	Reference	Reference	Reference
	Surgery recommended but not performed	0.85 (0.79–0.92)	0.87 (0.74–1.03)	0.85 (0.78–0.92)
	Surgery performed	0.70 (0.67–0.74)	0.61 (0.55–0.68)	0.72 (0.69–0.76)
Year of cancer diagnosis	2000–2001	Reference	Reference	Reference
	2002–2003	1.06 (0.98–1.14)	1.12 (0.94–1.32)	1.04 (0.96–1.13)
	2004–2005	1.03 (0.95–1.11)	0.77 (0.58–1.01)	1.05 (0.96–1.14)
	2006–2007	0.94 (0.87–1.01)	0.72 (0.55–0.95)	0.96 (0.88–1.05)
	2008–2009	0.94 (0.87–1.01)	0.78 (0.59–1.02)	0.93 (0.86–1.01)
	2010–2011	0.88 (0.81–0.95)	0.76 (0.58–1.00)	0.87 (0.80–0.95)
	2012–2013	0.87 (0.81–0.95)	0.74 (0.56–0.98)	0.87 (0.80–0.95)
	2014–2015	0.84 (0.77–0.91)	0.71 (0.54–0.94)	0.84 (0.77–0.92)
	2016–2017	0.71 (0.65–0.79)	0.68 (0.51–0.91)	0.72 (0.65–0.81)

surgery was also statistically significantly associated with survival.[10] Another study using more than 35 years of SEER data found younger age, early stage, and treatment with surgery as independent factors associated with longer survival.[27] Similar findings were reported from analyses of data from the National Cancer Database, where better survival was seen among patients of younger age, female sex, and epithelioid histology.[28] Different from these previous studies, our RSF-based analysis did not find sex and race/ethnicity to be highly predictive of survival, and stage was an important predictor only for female patients. One explanation is that, in the traditional CPH model, the significant HRs reflect the relative comparisons of the risk of death within a specific covariate (eg, mortality risk is higher in males than females), whereas RSF identifies how well a variable predicts mortality (eg, sex is not an important variable to predict death, as all cases have high risk of dying). It is possible that the lack of variation in sex and race/ethnicity (approximately 80% of mesothelioma cases were male and non-Hispanic White) makes them less useful in predicting survival in RSF. Other potential explanations include differences in study designs, such as different data collection periods, patient selection criteria, and covariates

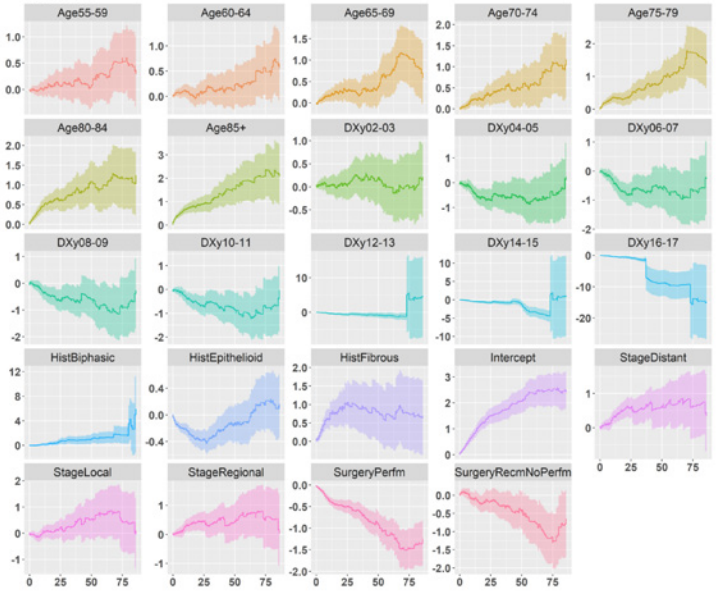
Figure 2. Time-Varying Effects of the Top 5 Covariates on Survival among (a) Overall, (b) Female, and (c) Male Patients

The X-axis indicates survival time (months). The Y-axis indicates β coefficients (ie, slopes) and their 95% CIs (shaded band) of the Aalen's additive regression model. The model used the top 5 predictors of survival selected from the random survival forest (RSF) machine-learning method. A visualization of the slopes for individual predictors over time helps to illustrate the time dependence of these cumulative covariate effects, and serves as a graphical examination of the proportional hazards (PH) assumption in the Cox Proportional Hazard
(CPH) model.

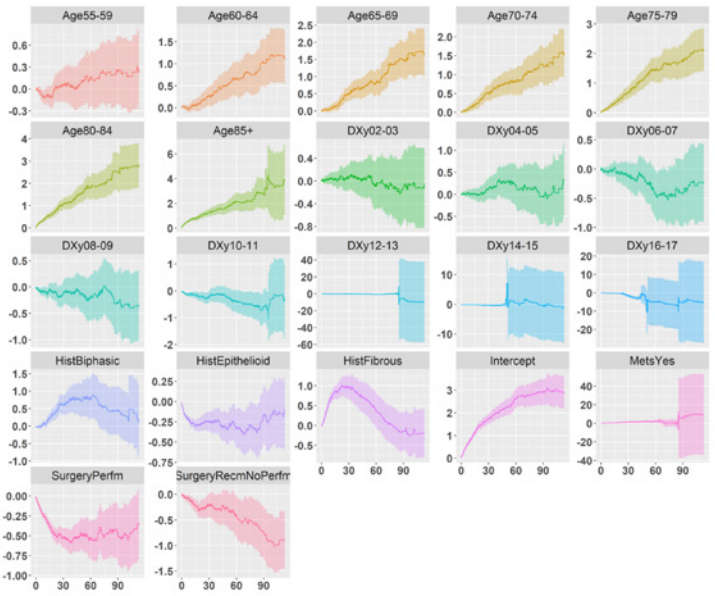
a. Overall



b. Female



c. Male



adjusted in the models.

While both RSF and CPH models can give information with regard to variable importance ranking, the capability of RSF in handling complex nonlinear relationship among large multidimensional data is likely to yield a more robust predication than CPH models. The relative importance ranking of covariates in predicting MPM mortality varied between male and female patients. Except for age, which was the most important predictor, the relative ranking of the other covariates varied considerably among the 3 patient samples used. For example, histological type ranked higher than surgery status among males than among females. Urban/rural status and median household income ranked fifth and seventh in the overall patients, respectively; fifth and sixth among males; but only seventh and ninth among females. While race/ethnicity was the fifth most important predictor of female survival, it reduced prediction (ie, negative VIMP) among males and overall patients. These differences may reflect the underlying discrepancies between male and female MPM patients. Sex differences in mesothelioma survival have long been noted. For example, one study found nearly a 10-point survival advantage among women.[29] Potential explanations of these discrepancies include the extent of exposure, disease status, and other underlying health conditions.

The observed consistency in influential variables of survival found between RSF and previous studies was reassuring. This is consistent with findings that have been reported for other oncological studies.[14-17] The similar prediction errors and IBS among the Kaplan–Meier, CPH, and RSF models were unexpected, as we anticipated better prediction accuracy from RSF. One explanation is the high proportion of deaths in the sample. When the prevalence of death is high (or survival is rare), predicting all patients are diseased (or no one survived) offers high accuracy that is not so useful for prognosis.[30] Another possible explanation may be the limited number of potential candidate predictors used in the model. Only 1 out of 10 candidate variables was found to be unimportant in predicting survival in the overall data, while none of the 9 candidate variables were excluded in male and female subgroups. While the results suggest that machine-learning methods yield results as good as classical methods, RSF may still be advantageous, as the proportional hazard assumption in CPH was not met in the current data, while RSF is not bound by such

restrictive assumptions.

This study had a few limitations. Due to data availability, our analysis did not include some important factors known to affect MPM survival, such as patient comorbidities and information about other therapy such as chemotherapy, radiation, or multimodal therapy.[28,31,32] In addition, detailed surgery types, which have been found to affect survival, were not available.[33,34] Similarly, we were only able to adjust for 2 area-level socioeconomic variables, and information regarding the extent of asbestos exposure and smoking behavior are not available in SEER data. The pending release of a research-enhanced version of SEER data, which will include additional cancer and treatment information, [20] will allow us to extend the current study.

Conclusion

We demonstrated the utility of RSF machine-learning algorithms in identifying the relative importance of factors associated with survival among MPM patients based on the most up-to-date mesothelioma data from SEER. RSF is a feasible complement to the commonly used CPH model, and a viable alternative, particularly when the proportional hazard assumption is unmet. When combined with traditional CPH approaches, RSF can be a powerful tool to help understand and quantify risk factors in predicting cancer survival. Additionally, RSF identified discrepancies between males and females in the importance ranking of covariates, which may help explain the sex differences in MPM survival.

References

1. Carbone M, Yang H. Mesothelioma: recent highlights. *Ann Transl Med*. 2017;5(11):238.
2. Ray M, Kindler HL. Malignant pleural mesothelioma: an update on biomarkers and treatment. *Chest*. 2009;136(3):888-896.
3. Liu B, van Gerwen M, Bonassi S, Taioli E. Epidemiology of environmental exposure and malignant mesothelioma. *J Thorac Oncol*. 2017;12(7):1031-1045.
4. Wagner JC, Sleggs CA, Marchand P. Diffuse pleural mesothelioma and asbestos exposure in the North Western Cape Province. *Br J Ind Med*. 1960;17:260-271.
5. PDQ Adult Treatment Editorial Board. Malignant mesothelioma treatment (adult) (PDQ): health professional version. National Cancer Institute website. Accessed June 23, 2020. **https://www.cancer.gov/types/mesothelioma/hp/mesothelioma-treatment-pdq**
6. Enewold L, Sharon E, Thomas A. Patterns of care and survival among patients with malignant mesothelioma in the United States. *Lung Cancer.* 2017;112:102-108.
7. Alpert N, van Gerwen M, Flores R, Taioli E. Gender differences in outcomes of patients with mesothelioma. *Am J Clin Oncol*. 2020;43(11):792-797.
8. Van Gerwen M, Alpert N, Wolf A, et al. Prognostic factors of survival in patients with malignant pleural mesothelioma: an analysis of the National Cancer Database. *Carcinogenesis*. 2019;40(4):529-536.
9. Thomas A, Chen Y, Yu T, Gill A, Prasad V. Distinctive clinical characteristics of malignant mesothelioma in young patients. *Oncotarget*. 2015;6(18):16766-16773.
10. Flores RM, Riedel E, Donington JS, et al. Frequency of use and predictors of cancer-directed surgery in the management of malignant pleural mesothelioma in a community-based (Surveillance, Epidemiology, and End Results [SEER]) population. *J Thorac Oncol*. 2010;5(10):1649-1654.
11. Beebe-Dimmer JL, Fryzek JP, Yee CL, et al. Mesothelioma in the United States: a Surveillance, Epidemiology, and End Results (SEER)-Medicare investigation of treatment patterns and overall survival. *Clin Epidemiol*. 2016;8:743-750.
12. Spirtas R, Connelly RR, Tucker MA. Survival patterns for malignant mesothelioma: the SEER experience. *Int J Cancer.* 1988;41(4):525-530.
13. Wang P, Li Y, Reddy CK. Machine learning for survival analysis: a survey. *ACM Comput Surv.* 2019;51(6).
14. Nicolo C, Perier C, Prague M, et al. Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer. *JCO Clin Cancer Inform*. 2020;4:259-274.
15. Omurlu IK, Ture M, Tokatli F. The comparisons of random survival forests and Cox regression analysis with simulation and an application related to breast cancer. *Expert Syst Appl*. 2009;36(4):8582-8588.
16. Datema FR, Moya A, Krause P, et al. Novel head and neck cancer survival analysis approach: random survival forests versus Cox proportional hazards regression. *Head Neck*. 2012;34(1):50-58.
17. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Sci Rep.* 2021;11(1):6968.
18. National Cancer Institute. Surveillance, Epidemiology, and End Results Program (SEER).
19. SEER. Incidence - SEER Research Data, 18 Registries, Nov 2019 Sub (2000–2017) - Linked To County Attributes - Time Dependent (1990-2017) Income/Rurality, 1969-2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2020, based on the November 2019 submission.
20. SEER data change history: April 2020 release. National Cancer Institute Surveillance, Epidemiology, and End Results Program website. **https://seer.cancer.gov/data/2020-data-changes.html**
21. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3):841-860.
22. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32.
23. Ishwaran H, Kogalur UB. Package 'randomForestSRC': Fast unified random forests for survival, regression, and classification (RF-SRC). The Comprehensive R Archive Network website. Updated 2020. **https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf**
24. Aalen OO, Scheike TH. Aalen's additive regression model. In: Armitage P, Colton T, eds. *Encyclopedia of Biostatistics*. Wiley; 2005.
25. Gerds TA, Andersen PK, Kattan MW. Calibration plots for risk prediction models in the presence of competing risks. *Stat Med.* 2014;33(18):3191-3203.
26. Meyerhoff RR, Yang C-FJ, Speicher PJ, et al. Impact of mesothelioma histologic subtype on outcomes in the Surveillance, Epidemiology, and End Results database. *J Surg Res.* 2015;196(1):23-32.
27. Taioli E, Wolf AS, Camacho-Rivera M, et al. Determinants of survival in malignant pleural mesothelioma: a Surveillance, Epidemiology, and End Results (SEER) study of 14,228 patients. *PLoS One*. 2015;10(12):e0145039.
28. Keshava HB, Tang A, Siddiqui HU, et al. Largely unchanged annual incidence and overall survival of pleural mesothelioma in the USA. *World J Surg*. 2019;43(12):3239-3247.
29. Taioli E, Wolf AS, Camacho-Rivera M, Flores RM. Women with malignant pleural mesothelioma have a threefold better survival rate than men. *Ann Thorac Surg*. 2014;98(3):1020-1024.
30. Lever J, Krzywinski M, Altman N. Classification evaluation. *Nat Methods*. 2016;13(8):603-604.
31. Cao C, Tian D, Park J, Allan J, Pataky KA, Yan TD. A systematic review and meta-analysis of surgical treatments for malignant pleural mesothelioma. *Lung Cancer.* 2014;83(2):240-245.
32. Abdel-Rahman O, Elsayed Z, Mohamed H, Eltobgy M. Radical multimodality therapy for malignant pleural mesothelioma. *Cochrane Database Syst Rev*. 2018;1(1):CD012605.
33. Taioli E, Wolf AS, Flores RM. Meta-analysis of survival after pleurectomy decortication versus extrapleural pneumonectomy in mesothelioma. *Ann Thorac Surg*. 2015;99(2):472-480.

Original Article

Health Care Utilization Prior to Ovarian Cancer Diagnosis in Publicly Insured Individuals in New York State

Margaret Gates Kuliszewski, ScD[a, b], Francis P. Boscoe, PhD[b]; Victoria L. Wagner, MS[c]; Maria J. Schymura, PhD[a, b]

Abstract: Background: Women with early-stage ovarian cancer may be asymptomatic or present with nonspecific symptoms. We examined health care utilization prior to ovarian cancer diagnosis to assess whether women with higher utilization differed in their prognosis and outcomes compared to women with low utilization. Methods: Using Medicaid, Medicare, and New York State Cancer Registry data for ovarian cancer cases diagnosed in 2006–2015, we examined selected health care visits that occurred 1–6 months before ovarian cancer diagnosis. We used multivariable-adjusted logistic regression to estimate odds ratios (ORs) and 95% CIs for associations of sociodemographic factors with number of prediagnostic visits and number of visits with tumor characteristics, and Cox proportional hazards regression to examine differences in survival by number of visits. Results: Women with >5 vs 0 prediagnostic visits were statistically significantly less likely to be diagnosed with distant vs local stage disease (OR, 0.72; 95% CI, 0.54–0.96), and women with 3–5 or >5 vs 0 prediagnostic visits had better overall survival (hazard ratio [HR], 0.88; 95% CI, 0.80–0.96 and HR, 0.90; 95% CI, 0.83–0.98, respectively). In stratified analyses, the association with improved survival was observed only among cases with regional or distant stage disease. Conclusions: Women with high health care utilization prior to ovarian cancer diagnosis may have better prognosis and survival, possibly because of earlier detection or better access to care throughout treatment. Women and their health care providers should not ignore symptoms potentially indicative of ovarian cancer and should be persistent in following up on symptoms that do not resolve.

Key words: *cancer registry, health care utilization, Medicaid, Medicare, ovarian cancer, survival, symptoms*

Introduction

Each year, approximately 313,959 cases of ovarian cancer are diagnosed and 207,252 women die from ovarian cancer worldwide,[1] with an estimated 21,410 new cases and 13,770 deaths in the United States.[2] Although ovarian cancer is less common than other cancers affecting women, the majority of cases are diagnosed at an advanced stage when prognosis is poor. In the United States, 16% of cases are localized and 21% are regional at diagnosis, and the corresponding 5-year relative survival rates are 93% and 75%, respectively. However, almost two-thirds of women are diagnosed with distant (56%) or unstaged (7%) disease, and only approximately 30% of these women survive for at least 5 years after diagnosis.[3] The high proportion of cases diagnosed with advanced disease is primarily due to the lack of a suitable screening tool for ovarian cancer and the absence of easily recognizable symptoms of early-stage disease.

Previous studies have documented the presence of nonspecific symptoms in women with early-stage ovarian cancer, including abdominal, pelvic, or lower back pain; abdominal or pelvic bloating/fullness; fatigue; difficulty eating or early satiety; urinary frequency or urgency; and constipation.[4-10] However, these symptoms may be overlooked or misdiagnosed as other conditions. In a retrospective study of symptoms in ovarian cancer cases and population-based controls, Rossing et al reported that using symptoms to identify women who should be screened for ovarian cancer would only detect 1 case in 100 women with such symptoms and diagnosis might not occur early enough to improve prognosis.[9] Similarly, Lim et al found that symptoms often began within 3 months of diagnosis, when screening for symptoms is unlikely to markedly accelerate diagnosis.[11] However, other studies suggest that screening symptomatic women for ovarian cancer may have benefits, including earlier diagnosis and improved survival.[7,12]

Although women with ovarian cancer often experience symptoms of their disease before diagnosis, it is unknown whether differences in health care utilization because of these symptoms influence time of detection and therefore patient prognosis. Using data from the New York State Cancer Registry (NYSCR), Medicaid, and fee-for-service Medicare, we estimated the number of medical visits where we hypothesized there was an opportunity for detection of an ovarian mass or where symptoms might have generated concern and resulted in additional follow-up. We then examined the number of prediagnostic visits in relation to sociodemographic factors, tumor characteristics, and

[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. [c] Office of Quality and Patient Safety, New York State Department of Health, Albany, 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 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 opinions, findings, and conclusions expressed here are those of the authors and do not necessarily represent the official views of the funding agencies.

overall survival to determine whether women with a larger number of visits differed in their characteristics, prognosis, and outcomes compared to women with fewer prediagnostic visits.

Methods

Study Design

Our study population included incident cases of epithelial ovarian cancer that were diagnosed in 2006–2015, reported to the NYSCR, and enrolled in fee-for-service Medicare or Medicaid for at least 6 months before and after diagnosis with no more than 1 month of not being enrolled. We restricted our analysis to this time period because of the unavailability of complete claims data before 2005 (Medicaid) and after 2016 (Medicare).

We identified 15,387 invasive ovarian cancer cases diagnosed in New York State (NYS) in 2006–2015. Of these, we sequentially excluded 2,110 subsequent primaries, 1,550 nonepithelial tumors, 6 cases diagnosed based on autopsy only, 111 cases who survived less than 7 days after diagnosis (since these cases may differ from the larger population of cases in NYS), 225 cases with missing month of diagnosis, and 10 cases who were younger than 18 years (n = 8) or older than 100 years (n = 1) at diagnosis or who did not identify as female (n = 1). Of the remaining 11,375 cases, we excluded 6,269 patients who did not meet the inclusion criteria for being publicly insured and continuously enrolled, or who were enrolled in Medicare Managed Care and did not have claims data available. After these exclusions our analysis included 5,106 cases, including 717 cases who were continuously enrolled in Medicaid, 3,808 cases who were enrolled in fee-for-service Medicare and had at least 1 prior carrier or outpatient claim, and 581 cases who met all of these criteria and were considered dually enrolled. This analysis was approved by the Institutional Review Board of the NYS Department of Health.

Assessment of Health Care Utilization

We identified health care visits that occurred 1–6 months before diagnosis for each case. We included specific Current Procedural Terminology (CPT) procedure codes and International Classification of Diseases, Ninth (ICD-9) and Tenth Revision (ICD-10) diagnosis codes for general office or preventive care visits, nonmalignant and noninfectious gynecological conditions, and possible symptoms of ovarian cancer. Appendix 1 lists CPT, ICD-9, and ICD-10 codes included as health care visits in this analysis. We excluded visits with a missing service date and visits that occurred after diagnosis or less than 30 days before diagnosis to avoid including visits related to diagnostic services.

For each case, we determined the total number of relevant health care visits and claims that occurred 1–6 months (30–180 days) before diagnosis. We then created categorical variables based on approximate quartiles for visits (0, 1–2, 3–5, or >5 visits) and claims 1–6 months before diagnosis (0, 1–3, 4–10, or >10 claims). We secondarily examined associations with number of visits and claims 1–12 months before diagnosis.

Ovarian Cancer Characteristics and Survival

Ovarian cancer cases diagnosed in NYS were retrieved from the NYSCR SEER*DMS database using the Surveillance, Epidemiology, and End Results Program (SEER) Site Recode ICD-O-3 value of 27040. Noninvasive and nonepithelial cancers were excluded using the SEER variables Behavior Recode for Analysis and ICD-O-3 Histology. Cases were categorized by summary stage (local, regional, distant, unknown), grade (well/moderately well differentiated, poorly differentiated/undifferentiated, unknown), and histologic subtype (serous, mucinous, endometrioid, other/mixed epithelial histology). Survival analyses used a SEER variable for number of months between date of ovarian cancer diagnosis and date of last known active follow up or date of death, with a censoring date of December 31, 2018. Secondary analyses used an alternate SEER survival variable where patients without a known date of death were presumed to be alive as of the censoring date. Mortality data were obtained from the National Death Index and the Vital Records Offices of the NYS Department of Health and the New York City (NYC) Department of Health and Mental Hygiene.

Covariates

Additional data on patient characteristics were retrieved from the NYSCR SEER*DMS database. Age at diagnosis was categorized in approximate quintiles (18–64, 65–69, 70–74, 75–79, 80–100 years) or modeled using a restricted cubic spline to allow for a nonlinear relationship. Race was categorized as White, Black, or other, and Hispanic ethnicity was categorized as Hispanic or non-Hispanic. Marital status at diagnosis was categorized as single, married, divorced/separated, widowed, or other/unknown. Insurance enrollment was determined based on Medicaid and Medicare data and categorized as Medicaid only, Medicare only, or both Medicaid and Medicare (dual enrollment). Location of residence at diagnosis was categorized as NYC/Long Island or rest of state. Data on receipt of surgery or chemotherapy during the patient's initial treatment were categorized as yes, no, or unknown.

Statistical Analysis

We examined descriptive characteristics of cases overall and by number of visits 1–6 months before diagnosis. We calculated *P* values for the association between each descriptive characteristic and categories for number of health care visits using χ^2 tests for categorical variables and the Kruskal-Wallis test for continuous variables. We used multinomial unconditional logistic regression to calculate multivariable-adjusted odds ratios (ORs) and 95% CIs for associations of sociodemographic characteristics of interest, including age, race/ethnicity, insurance enrollment, location of residence, and marital status, with number of prediagnostic visits. Because of zero cell counts for the combination of older age categories and insurance enrollment in Medicaid only, we adjusted analyses of categorical age and number of prediagnostic visits for insurance enrollment using a binary variable (Medicare only vs all others). For all other analyses of predictors of number of

prediagnostic visits, we adjusted for age using a restricted cubic spline with 3 knots or turning points based on percentiles of the data. This was because of evidence of a nonlinear association with age.

Next, we examined the number of prediagnostic visits as the exposure variable in relation to tumor characteristics and overall survival through December 31, 2018. We used binary and multinomial unconditional logistic regression to calculate multivariable-adjusted ORs and 95% CIs for the association of the number of prediagnostic visits with disease stage, differentiation, and histology, and Cox proportional hazards regression to calculate multivariable-adjusted hazard ratios (HRs) and 95% CIs for the association of number of prediagnostic visits with overall survival and ovarian cancer-specific survival. In addition, we used Kaplan–Meier curves to evaluate differences in survival distributions by number of prediagnostic visits, and tested for differences in survival distributions using the score test. For each model, we used a stepwise backward approach to select the final model, starting with the full model adjusted for age using a restricted cubic spline and categories for all other covariates. We removed each covariate 1 at a time starting with the variable with the largest nonsignificant *P* value for the association with the outcome and examined the change in estimate after removal of the variable. Covariates that did not change any estimate by more than 5% were dropped from the model as a conservative estimate of confounding. Age, disease stage, tumor differentiation (for logistic regression models only), and histology were retained in the models regardless of their status as confounders.

We evaluated potential effect modification of the association between number of prediagnostic visits and tumor characteristics or overall survival by modeling interaction terms between the exposure and modifier of interest and conducting likelihood ratio tests comparing models with and without interaction terms. We evaluated categories of age, race, insurance enrollment, location of residence, disease stage, tumor differentiation, and histology as potential effect modifiers.

In secondary analyses, we excluded outlying values for number of prediagnostic visits, where outliers were defined as values more than 3 standard deviations from the mean value. Finally, we restricted the analysis to prediagnostic visits where 1 or more symptoms of ovarian cancer (Appendix 2) was diagnosed, to assess whether the associations differed when only visits with reported ovarian cancer symptoms were considered. All analyses were conducted using SAS version 9.4 (SAS Institute Inc).

Results

On average, the 5,106 women with ovarian cancer in our sample were aged 71 years with 5 relevant health care visits 1–6 months before diagnosis (Table 1). The majority of women in our sample were White (83%), non-Hispanic (90%), married or widowed (67%), and enrolled in Medicare only (75%). The stage distribution was 10% localized, 16% regional, and 69% distant, with the remainder unknown, similar to the distribution reported by SEER for women aged 65 years and older.[3] χ^2 Tests indicated that women with a higher number of prediagnostic visits were less likely to be Black and were more likely to be enrolled in Medicaid only or dually enrolled, residents of NYC or Long Island, and to have localized or regional disease (Table 2).

Table 1. Characteristics of 5,106 Publicly Insured Ovarian Cancer Cases Diagnosed in 2006–2015 in New York State

Characteristic	*Mean (SD) or n (%)*
Mean age in years	71.4 (12.6)
Mean number of health care visits 1–6 months before diagnosis*	4.7 (7.3)
Insurance enrollment, n (%)	
Medicaid	717 (14.0)
Medicare	3,808 (74.6)
Both Medicaid and Medicare	581 (11.4)
Race, n (%)	
White	4,225 (82.8)
Black	616 (12.1)
Other	265 (5.2)
Hispanic ethnicity, n (%)	523 (10.2)
Marital status, n (%)	
Single	1,020 (20.0)
Married	1,888 (37.0)
Divorced or separated	528 (10.3)
Widowed	1,544 (30.2)
Other or unknown	126 (2.5)
Residence in New York City or Long Island, n (%)	2,748 (53.8)
Disease stage at diagnosis, n (%)	
Local	492 (9.6)
Regional	811 (15.9)
Distant	3,519 (68.9)
Unknown	284 (5.6)
Tumor differentiation, n (%)	
Well/moderately well differentiated	659 (12.9)
Poorly/undifferentiated	2,402 (47.0)
Unknown	2,045 (40.1)
Tumor histology, n (%)	
Serous	2,568 (50.3)
Mucinous	240 (4.7)
Endometrioid	307 (6.0)
Other or mixed epithelial	1,991 (39.0)
Received surgery during initial treatment	3,400 (66.6)
Received chemotherapy during initial treatment	3,247 (63.6)

Note: Percentages may not sum to 100 due to rounding.

*Restricted to general office or preventive care visits, visits for nonmalignant and noninfectious gynecological conditions, and visits possibly related to symptoms of ovarian cancer.

In multivariable-adjusted predictive analyses of number of visits 1–6 months before diagnosis, age, race, ethnicity, insurance enrollment, and location of residence were statistically significant predictors (Table 3), whereas marital status was unassociated with number of prediagnostic visits (results not shown). Compared to women aged 18–64 years, women in all other age groups were more likely to have >5 vs 0 relevant health care visits. Women enrolled in Medicaid only or dually enrolled were statistically significantly more likely to have 1–2, 3–5, or >5 vs 0 prediagnostic visits, compared to women enrolled in Medicare only. In contrast, Black women and Hispanic women were

Table 2. Characteristics of 5,106 Publicly Insured Ovarian Cancer Cases Diagnosed in 2006–2015 in New York State by Number of Health Care Visits 1–6 Months before Diagnosis

Characteristic	*Number of visits**				*P* †
	0	1–2	3–5	>5	
Number of women	1,421	935	1,110	1,640	
% of total	27.8	18.3	21.7	32.1	
Mean (SD) age in years	72 (11)	71 (13)	71 (14)	71 (13)	0.30
Insurance enrollment (%)					
Medicaid	7.5	14.6	17.9	16.8	<0.001
Medicare	89.8	75.4	70.2	63.9	
Both Medicaid and Medicare	2.7	10.1	11.9	19.3	
Race (%)					
White	81.4	84.6	82.0	83.4	<0.001
Black	15.2	11.1	10.7	10.8	
Other	3.4	4.3	7.3	5.9	
Hispanic ethnicity (%)	9.4	9.1	10.4	11.5	0.15
Marital status (%)					
Single	17.6	20.9	19.9	21.6	0.23
Married	39.0	35.4	36.6	36.4	
Divorced or separated	9.2	11.0	11.5	10.2	
Widowed	31.4	30.4	29.6	29.6	
Other or unknown	2.9	2.4	2.3	2.3	
Residence in New York City or Long Island (%)	50.4	48.5	54.0	59.8	<0.001
Disease stage at diagnosis (%)					
Local	8.2	8.3	10.2	11.2	<0.001
Regional	14.4	13.7	16.6	18.0	
Distant	71.9	73.3	67.9	64.6	
Unknown	5.6	4.7	5.3	6.2	
Tumor differentiation, n (%)					
Well/moderately well differentiated	10.9	12.8	13.4	14.3	0.006
Poorly/undifferentiated	46.4	44.5	47.6	48.7	
Unknown	42.7	42.7	39.0	37.0	
Tumor histology, n (%)					
Serous	48.0	51.2	52.3	50.4	0.005
Mucinous	4.4	4.1	4.4	5.5	
Endometrioid	5.0	4.8	6.5	7.3	
Other/mixed epithelial	42.6	39.9	36.8	36.9	

Note: Percentages may not sum to 100 due to rounding.

*Restricted to general office or preventive care visits, visits for nonmalignant and noninfectious gynecological conditions, and visits possibly related to symptoms of ovarian cancer.

†*P* value for unadjusted association from Kruskal-Wallis test for age and χ^2 test for categorical variables.

Table 3. Multivariable-Adjusted Odds Ratios (ORs) and 95% CIs for Sociodemographic Predictors of Number of Health Care Visits 1–6 months before Diagnosis for 5,106 Publicly Insured Ovarian Cancer Cases Diagnosed in 2006–2015 in New York State

Sociodemographic predictor	*Number of visits (outcome)‡*		
	1–2	**3–5**	**>5**
Age at diagnosis, y*			
18–64	1.00 (ref)	1.00 (ref)	1.00 (ref)
65–69	1.31 (0.93–1.84)	1.03 (0.74–1.42)	1.36 (1.00–1.85)
70–74	1.24 (0.88–1.75)	1.00 (0.72–1.39)	1.46 (1.07–1.98)
75–79	1.22 (0.86–1.75)	1.20 (0.85–1.67)	2.19 (1.60–2.99)
80–100	1.56 (1.12–2.18)	1.62 (1.18–2.21)	2.36 (1.75–3.18)
Race†			
White	1.00 (ref)	1.00 (ref)	1.00 (ref)
Black	0.58 (0.44–0.76)	0.52 (0.40–0.67)	0.44 (0.35–0.55)
Other	0.77 (0.48–1.21)	1.15 (0.77–1.72)	0.75 (0.51–1.12)
Ethnicity†			
Non-Hispanic	1.00 (ref)	1.00 (ref)	1.00 (ref)
Hispanic	0.71 (0.52–0.96)	0.72 (0.54–0.96)	0.65 (0.50–0.85)
Insurance enrollment†			
Medicare only	1.00 (ref)	1.00 (ref)	1.00 (ref)
Medicaid only	2.89 (1.88–4.44)	3.42 (2.28–5.14)	5.15 (3.51–7.56)
Both Medicaid and Medicare	5.24 (3.52–7.80)	6.47 (4.42–9.48)	12.38 (8.66–17.69)
Region†			
New York City or Long Island	1.00 (ref)	1.00 (ref)	1.00 (ref)
Rest of State	1.05 (0.88–1.26)	0.88 (0.74–1.05)	0.68 (0.58–0.80)

*Approximate quintiles of age at diagnosis. Estimates adjusted for all other covariates in the table with insurance enrollment collapsed to a binary variable for Medicare only vs all others.

†Estimates adjusted for all other covariates in the table with age modeled using a restricted cubic spline due to evidence of a nonlinear association.

‡Compared to 0 visits; restricted to general office or preventive care visits, visits for nonmalignant and noninfectious gynecological conditions, and visits possibly related to symptoms of ovarian cancer.

statistically significantly less likely to have 1–2, 3–5, or >5 vs 0 prediagnostic visits compared to White women, and women who lived outside of NYC or Long Island were statistically significantly less likely to have >5 vs 0 visits than women who resided in NYC or Long Island.

In analyses adjusted for age and other tumor characteristics, women with >5 vs 0 visits 1–6 months before diagnosis had 28% lower odds of being diagnosed with distant vs localized disease (OR, 0.72; 95% CI, 0.54–0.96) (Table 4). In addition, women with 3–5 or >5 vs 0 visits were statistically significantly less likely to be diagnosed with other/mixed histology tumors (OR, 0.74; 95% CI, 0.61–0.89 for 3–5 visits; OR, 0.79; 95% CI, 0.67–0.94 for >5 visits). No other clear associations between number of visits and disease stage, tumor differentiation, or histology were observed. The results were similar in secondary analyses of number of visits 1–12 months before diagnosis, but the association with distant vs localized disease was attenuated and no longer statistically significant (results not shown).

Women with a higher number of prediagnostic visits also had better overall survival (Table 4 and Figure 1). In multivariable-adjusted analyses, the hazard of death from any cause was 12% lower in women with 3–5 visits and 10% lower in women with >5 vs 0 visits 1–6 months before diagnosis (HR, 0.88; 95% CI, 0.80–0.96 for 3–5 visits; HR, 0.90; 95% CI, 0.83–0.98 for >5 visits). The results were unchanged when women without a known date of death were presumed to be alive as of the censoring date (results not shown). The results were also similar for analyses of ovarian cancer-specific survival (results not shown). The difference between the overall survival distribution functions for categories of prediagnostic visits 1–6 months before diagnosis was statistically significant ($P < .0001$). The survival distribution functions were nearly identical for women with 3–5 and >5 visits, and at every time point these women had better overall survival than those with 0 or 1–2 visits. In secondary analyses of the number of visits 1–12 months before diagnosis, the results were similar but slightly attenuated (results not shown).

For the associations presented in Table 4, there was a statistically significant interaction with disease stage for the association between number of visits 1–6 months before

Table 4. Multivariable-Adjusted Odds Ratios (ORs), Hazard Ratios (HRs), and 95% CIs for Associations of Number of Health Care Visits 1–6 Months before Diagnosis with Tumor Characteristics and Overall Survival for 5,106 Publicly Insured Ovarian Cancer Cases Diagnosed in 2006–2015 in New York State

*Outcome variable**	*Number of visits (exposure)§*			
	0	**1–2**	**3–5**	**>5**
Disease stage†				
Regional	1.00 (ref)	0.96 (0.66–1.41)	0.94 (0.66–1.32)	0.91 (0.67–1.24)
Distant	1.00 (ref)	1.09 (0.77–1.54)	0.84 (0.61–1.16)	0.72 (0.54–0.96)
Unknown	1.00 (ref)	0.83 (0.50–1.38)	0.85 (0.53–1.36)	0.96 (0.63–1.44)
Tumor differentiation†				
Poorly/ undifferentiated	1.00 (ref)	0.77 (0.56–1.05)	0.98 (0.73–1.33)	1.01 (0.77–1.32)
Unknown	1.00 (ref)	0.84 (0.60–1.16)	0.92 (0.67–1.26)	0.82 (0.62–1.09)
Tumor histology†				
Mucinous	1.00 (ref)	0.71 (0.45–1.13)	0.69 (0.45–1.07)	0.97 (0.67–1.41)
Endometrioid	1.00 (ref)	0.87 (0.55–1.36)	1.06 (0.71–1.59)	1.21 (0.84–1.73)
Other/mixed	1.00 (ref)	0.83 (0.68–1.01)	0.74 (0.61–0.89)	0.79 (0.67–0.94)
Overall survival‡				
HR (95% CI)	1.00 (ref)	1.00 (0.91–1.10)	0.88 (0.80–0.96)	0.90 (0.83–0.98)

*Reference categories for tumor characteristics are local stage disease, well/moderately well differentiated disease, and serous tumor histology.

†Estimates adjusted for age at diagnosis using a restricted cubic spline and the other tumor characteristics included in the table.

‡Survival analysis adjusted for age at diagnosis using a restricted cubic spline, insurance enrollment (Medicare only, Medicaid only, or both Medicaid and Medicare), disease stage, tumor histology, and receipt of surgery during initial treatment (yes, no, unknown).

§Restricted to general office or preventive care visits, visits for nonmalignant and noninfectious gynecological conditions, and visits possibly related to symptoms of ovarian cancer.

Figure 1. Overall Survival by Number of Visits 1–6 Months before Diagnosis for 5,106 Publicly Insured Ovarian Cancer Cases Diagnosed in 2006–2015 in New York State, Adjusted for Age, Insurance Enrollment, Disease Stage, Tumor Histology, and Receipt of Surgery during Initial Treatment (Score Test P Value < .0001)

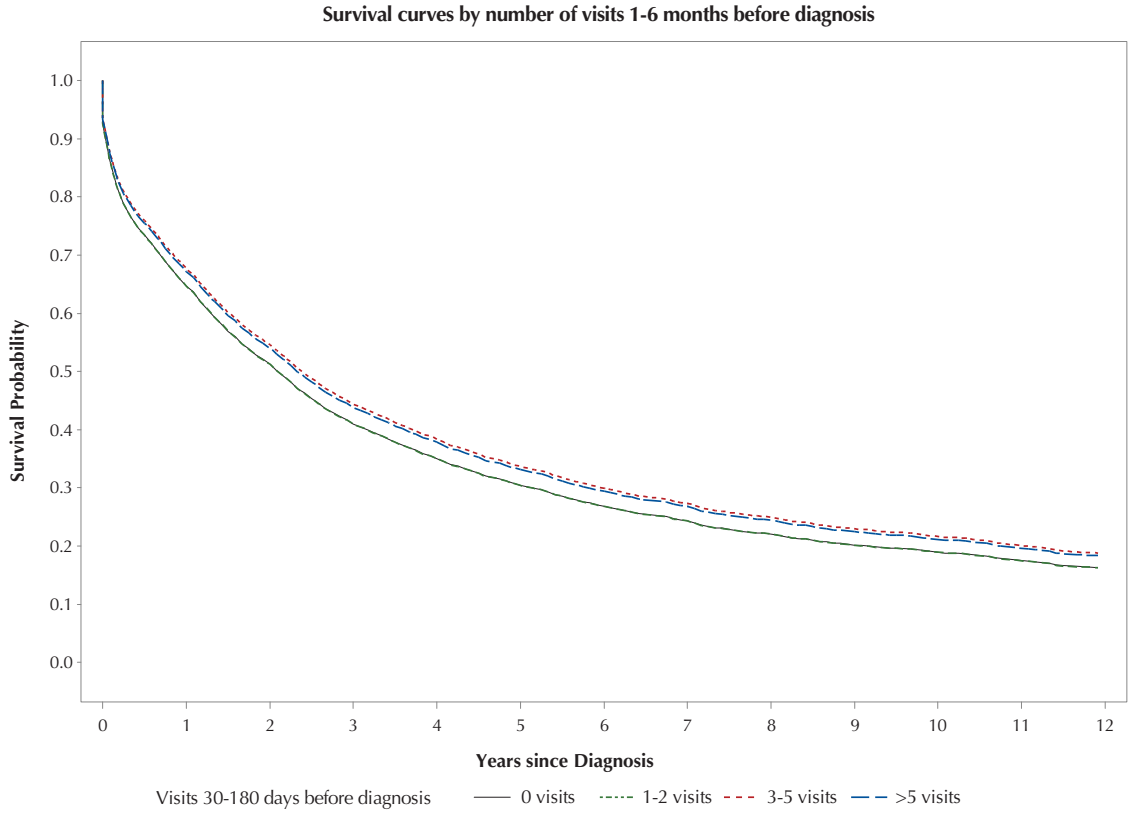


Table 5. Multivariable-Adjusted Hazard Ratios (HR) and 95% CIs for Associations of Number of Health Care Visits 1–6 Months before Diagnosis with Overall Survival, Stratified by Tumor Characteristics, for 5,106 Publicly Insured Ovarian Cancer Cases Diagnosed in 2006–2015 in New York State

Stratification variable	*Number of visits (exposure)*[†]				*P*[‡]
	0	1–2	3–5	>5	
Disease stage*					
Local	1.00 (ref)	2.28 (1.31–3.98)	1.21 (0.68–2.14)	1.96 (1.19–3.21)	.0006
Regional	1.00 (ref)	0.99 (0.73–1.32)	0.73 (0.55–0.98)	0.90 (0.70–1.16)	
Distant	1.00 (ref)	0.97 (0.87–1.07)	0.86 (0.77–0.95)	0.83 (0.75–0.92)	
Unknown	1.00 (ref)	0.99 (0.66–1.48)	1.23 (0.85–1.77)	1.25 (0.90–1.74)	
Differentiation*					
Well/moderate	1.00 (ref)	1.11 (0.79–1.57)	0.69 (0.48–1.00)	0.66 (0.47–0.91)	.09
Poorly/undifferentiated	1.00 (ref)	1.00 (0.86–1.15)	0.82 (0.71–0.94)	0.80 (0.70–0.91)	
Unknown	1.00 (ref)	0.99 (0.86–1.13)	0.96 (0.84–1.11)	1.05 (0.92–1.19)	
Histology*					
Serous	1.00 (ref)	1.07 (0.94–1.23)	0.91 (0.80–1.03)	0.88 (0.78–0.99)	.05
Mucinous	1.00 (ref)	2.08 (1.23–3.54)	1.11 (0.64–1.94)	1.23 (0.77–1.96)	
Endometrioid	1.00 (ref)	0.86 (0.50–1.49)	0.58 (0.33–1.03)	0.76 (0.46–1.25)	
Other/mixed	1.00 (ref)	0.92 (0.80–1.06)	0.86 (0.75–0.99)	0.91 (0.80–1.03)	

*Estimates adjusted for age at diagnosis using a restricted cubic spline, insurance enrollment (Medicare only, Medicaid only, or both Medicaid and Medicare), receipt of surgery during initial treatment (yes, no, unknown), and disease stage and/or tumor histology.

[†]Restricted to general office or preventive care visits, visits for nonmalignant and noninfectious gynecological conditions, and visits possibly related to symptoms of ovarian cancer.

[‡]*P* value for interaction from likelihood ratio test comparing models with and without interaction terms.

diagnosis and overall survival (*P* value for interaction = .0006). None of the other interactions examined were statistically significant at an α of 0.05, but there was evidence of several borderline statistically significant interactions including with both insurance enrollment and histology in relation to overall survival (both *P* values for interaction = .05). In survival analyses stratified by insurance enrollment, women enrolled in Medicaid only with 3–5 or >5 vs 0 visits before diagnosis had evidence of lower overall mortality (HR, 0.68; 95% CI, 0.48–0.96 and HR, 0.77; 95% CI, 0.56–1.07, respectively) (results not shown). In analyses stratified by disease stage, the association with improved overall survival was observed only among cases with regional or distant stage disease (Table 5). In contrast, among cases with localized disease there was evidence of poorer overall survival among women with 1 or more relevant health care visit 1–6 months before diagnosis. In analyses stratified by histologic subtype, the association between number of visits and improved survival was observed among women with serous tumors and those with other/mixed histology, but there were no clear associations between number of visits and survival for women with endometrioid or mucinous tumors.

The results presented in Tables 3 and 4 were essentially unchanged after excluding 43 women with outlying values for number of prediagnostic health care visits (≥27 relevant visits; results not shown). In addition, the results were similar when restricted to visits with a claim for 1 or more ovarian cancer symptoms listed in Appendix 2. Women in the highest category of symptom-related visits still had a statistically significant lower odds of distant stage disease, but the associations with overall survival were no longer statistically significant (results not shown).

Finally, we examined the most frequent symptom-related ICD-9/ICD-10 codes for visits that occurred 1–6 months before diagnosis to determine what symptoms of ovarian cancer were most commonly reported. On average, women had 1.8 visits with a claim for 1 or more ovarian cancer symptom (results not shown). Abdominal pain was the most common symptom, accounting for 24% of relevant claims for ovarian cancer symptoms, followed by gastroesophageal reflux (16%), abdominal or pelvic swelling, mass, or lump (12%), low or other back pain (11%), malaise and fatigue (10%), and constipation (7%). Of the 5,106 ovarian cancer cases included in the analysis, 1,077 (21%) had at least 1 claim for abdominal pain 1–6 months before diagnosis; 662 (13%) had claims for malaise and fatigue; 595 (12%) had claims for abdominal or pelvic swelling, mass, or lump; 485 (10%) had claims for gastroesophageal reflux; 419 (8%) had claims for low or other back pain; and 357 (7%) had claims for constipation. A total of 2,655 women (52%) did not have any clear symptom-related claims 1–6 months before diagnosis (results not shown).

Discussion

In this population of publicly insured women with invasive epithelial ovarian cancer, there was evidence that women with a higher number of selected health care visits

prior to diagnosis had better prognosis and survival when compared to women with lower health care utilization. These results suggest that, for some women, high health care utilization prior to ovarian cancer diagnosis is beneficial, possibly because it is a marker of better health care access throughout diagnosis and treatment.

To our knowledge, no prior studies have examined frequency of health care utilization immediately prior to ovarian cancer diagnosis in relation to disease prognosis and survival. In a previous analysis that linked SEER and Medicare data, Gornick et al reported no clear relationship between preventive service use and ovarian cancer stage at diagnosis. However, this study examined use of specific preventive services related to immunization and cancer screening, which may not reflect all possible opportunities for earlier diagnosis.[13]

In our analysis, we included specific encounters where we hypothesized there was an opportunity for detection of an ovarian mass or where symptoms may have led to additional testing. Previous studies have indicated that the majority of ovarian cancer patients are symptomatic before diagnosis. In a meta-analysis of symptoms prior to ovarian cancer diagnosis, 7.2% of women were asymptomatic when information was collected directly from the women and 22.6% were asymptomatic when symptoms were determined from hospital medical notes.[14] In a study of 124 women referred for suspected ovarian malignancy, all women later diagnosed with ovarian cancer were symptomatic before diagnosis but attributed the symptoms to normal changes related to aging, weight gain, or other natural processes.[4] In our sample, 52% of women did not have claims for any common symptoms of ovarian cancer prior to diagnosis, suggesting that some symptoms were not reported or were not captured in claims data. However, diagnostic and procedure codes for general medical and preventive care visits, which were included in our primary analyses, would be expected to capture some visits where ovarian cancer symptoms were reported but not recorded in claims. It is not possible with insurance claims data to identify ovarian cancer symptoms that were reported but not submitted for reimbursement purposes, likely leading to an underreporting of the prevalence of ovarian cancer symptoms in our data.

We observed that women with a greater number of prediagnostic visits were less likely to be diagnosed with distant stage disease. One possible explanation is that women who were persistent in seeking medical care for unresolved symptoms were more likely to be diagnosed at an earlier stage before the disease had metastasized. However, it is also possible that this result can be explained by differences in progression of certain ovarian tumors, with slower growing tumors allowing more time for diagnosis at an early stage, or differences in symptoms for different types of ovarian cancers where certain symptoms may be less likely to be misdiagnosed or more likely to prompt women to seek medical attention. Previous studies have reported that ovarian cancer symptoms may differ by disease stage and histologic subtype. In one study, women with serous cancers were more likely to report bowel symptoms, women with endometrioid cancers were more likely to report abnormal vaginal bleeding, and women with mucinous cancers were more likely to report abdominal distension. In addition, serous cancers were more likely to be diagnosed at an advanced stage and have a shorter duration of symptoms, while mucinous cancers were more likely to be diagnosed at an early stage and have a longer duration of symptoms.[15] Other studies have reported that women diagnosed with early-stage disease frequently report abdominal pain and/or distension,[7,16] and that these symptoms may be more likely to prompt women to seek medical attention than other common symptoms.[17] These studies suggest that certain symptoms may be more quickly reported to a health care provider and may often be characteristic of early-stage or less aggressive tumors.

We additionally observed that women with higher prediagnostic health care utilization had better survival, especially for cases with regional or distant stage disease and tumors with serous or other/mixed histology. This result could potentially be explained by better health care access throughout diagnosis and treatment, better ability to navigate the health care system, or better compliance with treatment. However, the fact that improvements in survival were observed only for cases with certain disease characteristics suggests the possibility that there were other contributing factors, such as earlier diagnosis within a specific category of disease stage. An analysis of symptomatic cases in an Australian case-control study indicated that decreasing the time from initial presentation or symptom onset to diagnosis did not have a statistically significant impact on survival.[18] However, limitations of the study design may have impacted the results, such as self-reported retrospective data on symptom onset and time to diagnosis. In contrast, in a retrospective cohort of Canadian women with invasive ovarian cancer, there was an association between time to diagnosis and overall survival, where survival decreased for delays greater than 80 days between initial presentation and diagnosis.[19] Although prior studies of the impact of diagnostic delays on survival are mixed, we cannot rule out the possibility that greater prediagnostic health care utilization resulted in improved survival for some cases by advancing the time of diagnosis.

This paper presents novel results from a large population of Medicaid and Medicare recipients in NYS. Our analysis included 5,106 invasive epithelial ovarian cancer cases with detailed diagnostic and claims information, allowing for analysis of specific procedure and diagnosis codes. The NYSCR has received gold-level certification since 1998 and routinely meets or exceeds all data standards for timeliness, completeness, and quality. Accurate mortality and survival data are obtained through linkages to other data sets, including NYS and NYC Vital Records. However, some limitations of the data may have affected the analysis results. Although accurate mortality data are available for US residents, we may have missed some deaths that occurred outside the United States. In addition, up-to-date follow-up information may not have been available for all cases, resulting in censoring as of the date of last known follow-up in survival analyses. Claims information was

Appendix 1. Current Procedural Terminology (CPT) and International Classification of Diseases, Ninth Revision (ICD-9) and Tenth Revision (ICD-10) Procedure and Diagnosis Codes Included as Prediagnostic Visits*	
CPT Code(s)	***Procedure***
99201–99205, 99211-99215	Office/other outpatient services
99241–99245	Office consultations
99381–99387, 99391-99397	Preventive medicine evaluation/reevaluation
G0101	Screening pelvic exam
G0123, G0124, G0141, G0143, G0144, G0145, G0147, G0148	Gynecologic cytopathology
G0344, G0402	Initial preventive physical examination
G0438-G0439	Annual well visit
G0463	Hospital outpatient clinic visit
G0467	Federally qualified health center visit, established patient
P3001	Screening Papanicolaou smear, cervical or vaginal
Q0091	Screening Papanicolaou smear
S0610, S0612	Annual gynecological examination
ICD-9 and ICD-10 Code(s)	***Diagnosis***
218, 2180-2182, 2189, D250-D252, D259	Uterine leiomyoma
219, 2191, 2198-2199, D261, D267, D269	Benign neoplasm of uterus
220, D279	Benign neoplasm of ovary
221, 2210-2212, 2218-2219, D280-D282, D287, D289	Benign neoplasm of other female genital organs
2190, D260	Benign neoplasm of cervix uteri
2360, D390	Neoplasm of uncertain behavior of uterus
2362, D3910	Neoplasm of uncertain behavior of ovary
2363, D398, D399	Neoplasm of uncertain behavior of other female genital organs
256, 2560-2564, 2568, 2569, 25631, 25639, E282, E28310, E28319, E2839, E288, E289	Ovarian dysfunction
53011, 53081, K210, K219	Gastro-esophageal reflux
5640, 56400-56401, 56409, K5900, K5901, K5904, K5909	Constipation
5693, K625	Bleeding, rectal
5781, K921	Blood in stool, melena
617, 6170-6174, 6178-6179, N800-N804, N808-N809	Endometriosis
619, 6190-6192, 6198-6199, N820, N824, N825, N828, N829	Fistula involving female genital tract
620, 6200, N8300-N8302	Follicular cyst of ovary
6201-6209, N8310-N8312, N8320, N83201, N83202, N83209, N8329, N83291, N83292, N83299, N83331, N83332, N83339, N8340-N8342, N8353, N837-N839	Other noninflammatory disorders of ovary, fallopian tube, and broad ligament
621, 6210-6219, 62130-62135, N840, N8500-N8502, N852-N859	Disorders of uterus not elsewhere classified
622, 6220-6229, 62210-62212, N841, N86, N870, N871, N879-N884, N888, N889	Noninflammatory disorders of cervix
6238, N898	Other specified noninflammatory disorders of vagina
6253, N946	Dysmenorrhea
6258	Other specified symptoms associated with female genital organs
6259	Female genital symptoms NOS
6260, N912	Absence of menstruation
6262, 6266, N920, N921, N924	Excessive or frequent menstruation

Appendix 1, cont. Current Procedural Terminology (CPT) and International Classification of Diseases, Ninth Revision (ICD-9) and Tenth Revision (ICD-10) Procedure and Diagnosis Codes Included as Prediagnostic Visits*	
ICD-9 and ICD-10 Code(s)	*Diagnosis*
6264, N925, N926	Irregular menstrual cycle
6268, 6269, N938, N939, N9489	Other or unspecified disorders of menstruation and other abnormal bleeding from female genital tract
6270	Premenopausal menorrhagia
6271, N950	Postmenopausal bleeding
N9983	Residual ovary syndrome
7242, M545	Low back pain
7245, M5489, M549	Other back pain
7807, 78079, R5381, R5383	Malaise and fatigue
78094, R6881	Early satiety
7830, R630	Anorexia (loss of appetite)
7832, 78321, 78322, R634, R636	Abnormal loss of weight, underweight
78701-78703, R110, R1111, R112	Nausea and/or vomiting
7871, R12	Heartburn
R140	Abdominal distension (gaseous)
7873, R141-R143	Flatulence, eructation, and gas pain
R194	Change in bowel habit
5645, 78791, K591, R197	Diarrhea
R198	Other specified symptoms and signs involving the digestive system and abdomen
78841, R350	Urinary frequency
78863, R3915	Urgency of urination
7890, 78900-78909, R1010-R1013, R1030-R1033, R1084, R109	Abdominal pain
R102	Pelvic and perineal pain
7893, 78930-78939, R1900-R1907, R1909	Abdominal or pelvic swelling, mass, or lump
7895, 78959, R188	Ascites
7896, 78960-78967, 78969, R10811-R10817, R10819, R10821-R10827, R10829	Abdominal tenderness
7899	Other symptoms involving abdomen and pelvis
7935	Nonspecific (abnormal) findings on radiological and other examination of genitourinary organs
7936, R935	Nonspecific (abnormal) findings on radiological and other examination of abdominal area
79500-79501, 79509, R87610, R87619	Abnormal Papanicolaou smear of cervix
79582, R971	Elevated CA-125
V700, V708, V709, Z0000, Z0001	General medical exam
V723, V7231, Z01411, Z01419	Gynecological exam
V7232	Encounter for Pap cervical smear to confirm findings of normal smear following abnormal smear
V762, Z124	Screening for malignant neoplasms of cervix
V7646, Z1273	Special screening for malignant neoplasms of ovary
V7647, Z1272	Special screening for malignant neoplasms of vagina

*Restricted to general office or preventive care visits, visits for nonmalignant and noninfectious gynecological conditions, and visits possibly related to symptoms of ovarian cancer.

Appendix 2. International Classification of Diseases, Ninth Revision (ICD-9) and Tenth Revision (ICD-10) Diagnosis Codes Included in Analyses of Prediagnostic Visits Where Symptoms of Ovarian Cancer Were Reported	
ICD-9 and ICD-10 Code(s)	***Diagnosis***
53011, 53081, K210, K219	Gastro-esophageal reflux
5640, 56400-56401, 56409, K5900, K5901, K5904, K5909	Constipation
5693, K625	Bleeding, rectal
5781, K921	Blood in stool, melena
6253, N946	Dysmenorrhea
6260, N912	Absence of menstruation
6262, 6266, N920, N921, N924	Excessive or frequent menstruation
6264, N925, N926	Irregular menstrual cycle
6268, 6269, N938, N939, N9489	Other or unspecified disorders of menstruation and other abnormal bleeding from female genital tract
6270	Premenopausal menorrhagia
6271, N950	Postmenopausal bleeding
7242, M545	Low back pain
7245, M5489, M549	Other back pain
7807, 78079, R5381, R5383	Malaise and fatigue
78094, R6881	Early satiety
7830, R630	Anorexia (loss of appetite)
7832, 78321, 78322, R634, R636	Abnormal loss of weight, underweight
78701-78703, R110, R1111, R112	Nausea and/or vomiting
7871, R12	Heartburn
R140	Abdominal distension (gaseous)
7873, R141-R143	Flatulence, eructation, and gas pain
R194	Change in bowel habit
5645, 78791, K591, R197	Diarrhea
78841, R350	Urinary frequency
78863, R3915	Urgency of urination
7890, 78900-78909, R1010-R1013, R1030-R1033, R1084, R109	Abdominal pain
R102	Pelvic and perineal pain
7893, 78930-78939, R1900-R1907, R1909	Abdominal or pelvic swelling, mass, or lump
7896, 78960-78967, 78969, R10811-R10817, R10819, R10821-R10827, R10829	Abdominal tenderness

not available for cases with private insurance or Medicare Managed Care coverage, and Medicaid cases without consistent enrollment were excluded from the analysis. As a result of these exclusions, cases included in the analysis were older on average and may not be representative of all invasive epithelial ovarian cancer cases in NYS. The data used in the analysis were collected for nonanalytic purposes and did not include several variables of interest, including detailed treatment data and other potential confounders. In addition, the claims data may not have captured all visits for ovarian cancer symptoms. However, the benefits of using claims data, including a reduced reliance on patient recall and the ability to retrieve relevant information on claims for a large number of cases, likely outweigh these limitations of the data.

Our results suggest that women with higher levels of preventive, gynecological, or symptom-related health care utilization immediately prior to ovarian cancer diagnosis have better prognosis and survival than women with lower health care utilization. The survival benefit appeared to be greatest for cases that typically have the worst prognosis, including regional and distant stage disease and tumors with serous or other/mixed histology. This benefit may be a result of earlier diagnosis or better access to health care throughout treatment. Confirmation of these results in another study population, particularly a representative population that includes privately insured individuals, is needed to assess generalizability of our results. Given the poor prognosis and survival for many women with ovarian cancer, these results have major implications and highlight the importance of access to care and persistence in following up on ovarian cancer symptoms that do not resolve. Women and their health care providers should be aware of the symptoms of ovarian cancer and should not ignore symptoms that may be indicative of ovarian cancer, especially when such symptoms occur in combination, persist, or worsen.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249.
2. American Cancer Society. Cancer Facts & Figures: 2021. American Cancer Society; 2021. **https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf**
3. SEER*Explorer. National Cancer Institute Surveillance, Epidemiology, and End Results Program website. Accessed October 22, 2021. **https://seer.cancer.gov/explorer/**
4. Bankhead CR, Collins C, Stokes-Lampard H, et al. Identifying symptoms of ovarian cancer: a qualitative and quantitative study. *BJOG*. 2008;115(8):1008-1014. doi:10.1111/j.1471-0528.2008.01772.x
5. Devlin SM, Diehr PH, Andersen MR, Goff BA, Tyree PT, Lafferty WE. Identification of ovarian cancer symptoms in health insurance claims data. *J Womens Health (Larchmt).* 2010;19(3):381-389. doi:10.1089/jwh.2009.1550
6. Goff BA, Mandel LS, Melancon CH, Muntz HG. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. *JAMA.* 2004;291(22):2705-2712. doi:10.1001/jama.291.22.2705
7. Lurie G, Thompson PJ, McDuffie KE, Carney ME, Goodman MT. Prediagnostic symptoms of ovarian carcinoma: a case-control study. *Gynecol Oncol.* 2009;114(2):231–236. doi:10.1016/j.ygyno.2009.05.001
8. Olson SH, Mignone L, Nakraseive C, Caputo TA, Barakat RR, Harlap S. Symptoms of ovarian cancer. *Obstet Gynecol.* 2001;98(2):212-217. doi:10.1016/S0029-7844(01)01457-0
9. Rossing MA, Wicklund KG, Cushing-Haugen KL, Weiss NS. Predictive value of symptoms for early detection of ovarian cancer. *J Natl Cancer Inst.* 2010;102(4):222-229. doi:10.1093/jnci/djp500
10. Ryerson AB, Eheman C, Burton J, et al. Symptoms, diagnoses, and time to key diagnostic procedures among older U.S. women with ovarian cancer. *Obstet Gynecol.* 2007;109(5):1053-1061. doi:10.1097/01.AOG.0000260392.70365.5e
11. Lim AWW, Mesher D, Gentry-Maharaj A, et al. Predictive value of symptoms for ovarian cancer: comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J Natl Cancer Inst.* 2012;104(2):114-124. doi:10.1093/jnci/djr486
12. Gilbert L, Basso O, Sampalis J, et al. Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. *Lancet Oncol.* 2012;13(3):285-291. doi:10.1016/S1470-2045(11)70333-3
13. Gornick ME, Eggers PW, Riley GF. Associations of race, education, and patterns of preventive service use with stage of cancer at time of diagnosis. *Health Serv Res.* 2004;39(5):1403-1427. doi:10.1111/j.1475-6773.2004.00296.x
14. Bankhead CR, Kehoe ST, Austoker J. Symptoms associated with diagnosis of ovarian cancer: a systematic review. *BJOG.* 2005;112(7):857-865. doi:10.1111/j.1471-0528.2005.00572.x
15. Lurie G, Wilkens LR, Thompson PJ, Matsuno RK, Carney ME, Goodman MT. Symptom presentation in invasive ovarian carcinoma by tumor histological type and grade in a multiethnic population: a case analysis. *Gynecol Oncol.* 2010;119(2):278-284. doi:10.1016/j.ygyno.2010.05.028
16. Webb PM, Purdie DM, Grover S, Jordan S, Dick ML, Green AC. Symptoms and diagnosis of borderline, early and advanced epithelial ovarian cancer. *Gynecol Oncol.* 2004;92(1):232-239. doi:10.1016/j.ygyno.2003.09.005
17. Smith EM, Anderson B. The effects of symptoms and delay in seeking diagnosis on stage of disease at diagnosis among women with cancers of the ovary. *Cancer.* 1985;56(11):2727-2732. doi:10.1002/1097-0142(19851201)56:11<2727::aid-cncr2820561138>3.0.co;2-8
18. Nagle CM, Francis JE, Nelson AE, et al. Reducing time to diagnosis does not improve outcomes for women with symptomatic ovarian cancer: a report from the Australian Ovarian Cancer Study Group. *J Clin Oncol.* 2011;29(16):2253-2258. doi:10.1200/JCO.2010.32.2164
19. Altman AD, Lambert P, Love AJ, et al. Examining the effects of time to diagnosis, income, symptoms, and incidental detection on overall survival in epithelial ovarian cancer: Manitoba Ovarian Cancer Outcomes (MOCO) Study Group. *Int J Gynecol Cancer.* 2017;27(8):1637-1644. doi:10.1097/IGC.0000000000001074

First Place Poster

Pilot Study on Early Onset of Colorectal Cancer in Patients Under Age 50: How and Why are they Diagnosed?

Bobbi Jo Matt, MS, RHIT, CTR[a]; Mary E. Charlton, PhD[b]; Richard Hoffman, MD, MPH[b]

Background

- Colorectal cancer is the fourth most commonly diagnosed cancer in the United States and the second leading cause of death.[1]
- Iowa's colorectal cancer incidence and mortality rates have decreased in those aged ≥ 50 years, while rates for those aged < 50 years have been increasing since 2000 (Figure 1).
- It is unclear if increases can be explained by increased high-risk screening, more diagnostic testing with colonoscopy, or changes in behavioral risk factors.

Study Aims

- Examine precipitating factors of colorectal cancer diagnosis in those aged < 50 years
- Determine the feasibility of collecting variables not routinely captured by cancer registries
- Create and pilot an abstraction form that can be used for future studies
- Determine the availability and feasibility of finding factor-specific variables in the central registry vs hospital records

Methods

Study Population

Inclusion criteria:

- Iowa residents ages 18 to 50 years
- Invasive, microscopically confirmed colorectal cancer diagnosed in 2017
- Colon (C180, C182–C187)
- Rectosigmoid junction (C199)
- Rectum (C209)
- Histologic types included in Colon & Rectal Cancer Collaborative Stage Schema ID: 00200, version 0204

Exclusion criteria:

- Diagnosed at autopsy, pathology, or death certificate only, and those identified only by recurrence/progression (nonanalytic cases)
- Carcinoid tumors and lymphomas

Study Design/Analysis

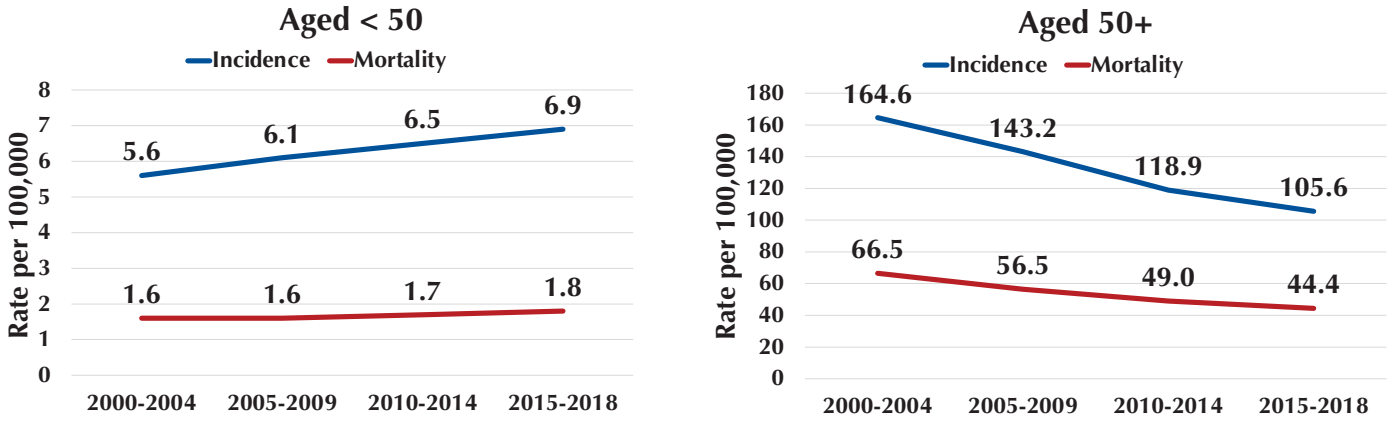
Study Design/Analysis

- Retrospective cross-sectional descriptive study
- Selected a sample of cases diagnosed in 2017 among those aged < 50 years, with an oversample of those aged < 40 years
- Trained registrars collected data from abstracts submitted to the Iowa Cancer Registry and hospital electronic health records where diagnostic services and/or treatment were received:
  - Reason(s) for seeking medical attention
  - Diagnostic testing
  - Risk factors
  - Staging
  - Molecular testing
- All analyses were conducted using SAS version 9.4

Key Findings

- In 95% of all cases, symptoms were the primary reason for seeking medical attention (Table 1, Figure 2).
- 33% of cases reported having a family history of colorectal polyps or a colorectal cancer.

Figure 1. Colorectal Age Adjusted Incidence[2] and Mortality[3] Rates by Age at Diagnosis, SEER 18, 2000-2018

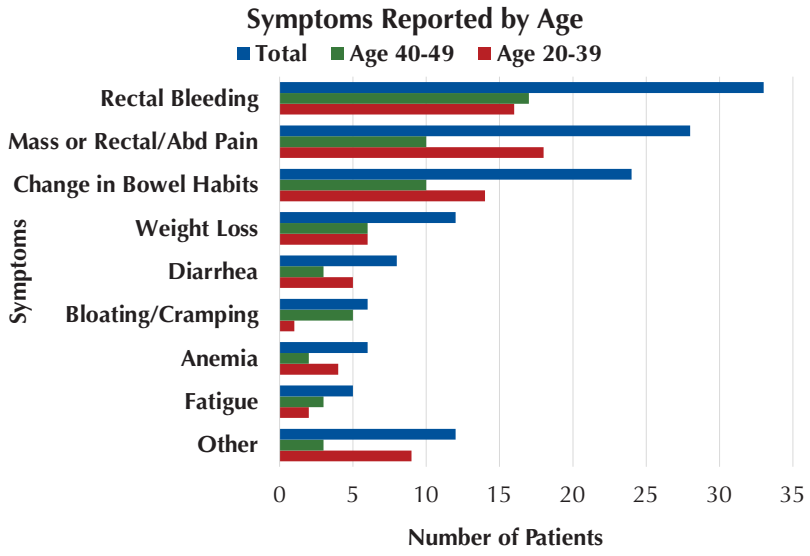


[a]Department of Epidemiology, University of Iowa College of Public Health and Iowa Cancer Registry. b Department of Internal Medicine, University of Iowa

Table 1. Patient and Tumor Characteristics by Age at Diagnosis			
Characteristics	*All, n (%) (n = 43)*	*Age 20–39 y, n (%) (n = 22)*	*Age 40–49 y, n (%) (n = 21)*
Race (White)	42 (98)	21 (95)	21 (100)
Ethnicity (non-Hispanic)	40 (93)	20 (95)	20 (95)
Marital status (married/domestic partner)	21 (49)	8 (36)	13 (62)
Residence (metropolitan)	28 (65)	14 (64)	14 (67)
Smoking status (current)	6 (14)	†	†
Alcohol status (current)	25 (58)	12 (55)	13 (62)
Body mass index (obese)	21 (50)	10 (45)	11 (55)
Reason for diagnosis (symptoms)	41* (95)	21 (95)	20 (95)
Family history (any)	24 (56)	13 (59)	11 (52)
Colorectal polyps	6 (14)	†	†
Colorectal cancer	10 (23)	6 (27)	†
Other cancer	16 (42)	8 (36)	8 (38)
High-risk comorbidities (any)	24 (56)	11 (50)	13 (62)
Obesity	21 (49)	10 (45)	11 (52)
Diabetes	†	†	†
Inflammatory bowel disease	†	†	†
Primary site			
Right (C180, C182, C183, C184)	7 (16)	†	†
Left (C185, C186, C187, C199)	18 (42)	10 (46)	8 (38)
Rectum (C209)	18 (42)	8 (36)	10 (48)
Staging (Summary Stage 2000)			
Localized	†	†	†
Regional	23 (53)	12 (54)	11 (52)
Distant	15 (35)	7 (32)	8 (38)
Genetic counseling (done)	14 (33)	7 (32)	7 (33)
Germline testing (done)	16 (37)	9 (41)	7 (33)

* Two were also high risk/surveillance. †Suppressed due to small numbers.
Fisher's exact tests were conducted for each characteristic and no statistically significant ($P < .05$) differences were found.

Figure 2. Common Symptoms Reported by Patient by Age at Diagnosis



- 56% of cases reported having a high-risk comorbid condition (diabetes, inflammatory bowel disease, and obesity).
- Comparisons between data collected from the Iowa Cancer Registry and hospital medical records demonstrated that patients' weight, height, alcohol and smoking status, family history, comorbidities, and genetic counseling were best identified from hospital record review, whereas staging, treatment and diagnostic testing could be identified in the registry's abstracts.

Summary and Conclusion

- Clinical symptoms were the predominant reason why those aged < 50 years sought medical care and presented with advanced stage.
- Developed and piloted the abstraction form for future studies and identified which variables could be found in the hospital records vs. the central registry's database.
- This work is an important step in informing a larger study with multivariable analysis to identify the primary factors driving the increasing incidence of colorectal cancer among those aged < 50 years.

Acknowledgments

Thank you to Michele West and Amanda Kahl for their assistance on this project.

References

1. Cancer stat facts: common cancer sites. National Cancer Institute Surveillance, Epidemiology, and End Results Program website. **https://seer.cancer.gov/statfacts/html/common.html**
2. Surveillance, Epidemiology, and End Results (SEER) Program (**www.seer.cancer.gov**) SEER*Stat Database: Incidence—SEER Research Plus Data, 18 Registries, Nov 2020 Sub (2000-2018)—Linked To County Attributes—Total U.S., 1969-2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission.
3. Surveillance, Epidemiology, and End Results (SEER) Program (**www.seer.cancer.gov**) SEER*Stat Database: Mortality—All COD, Aggregated With State, Total U.S. (1969–2018) <Katrina/Rita Population Adjustment>, National Cancer Institute, DCCPS, Surveillance Research Program, released May 2020. Underlying mortality data provided by NCHS (**www.cdc.gov/nchs**).

Second Place Poster

Do Modifiable Risk Factors Impact Pancreatic Cancer Survival in a Population-based Study after Adjusting for Clinical Factors?

Mei-Chin Hsieh, PhD, CTR[a]; Lu Zhang, PhD[b]; Cruz Velasco-Gonzalez, PhD[c]; Yong Yi, PhD[a]; Lisa A. Pareti, BS, CTR[a]; Edward J. Trapido, ScD, FACE[a]; Vivien W. Chen, PhD[a]; Xiao-Cheng Wu, MD, MPH, CTR[a]

Key Messages

- The impact of diabetes and smoking on cancer-specific (CS) survival was greater among patients with stage I–II than those with stage III–IV.
- Only a body mass index (BMI) ≥ 35 kg/m^2 was observed to increase risk of CS death among stage III–IV patients.
- As expected, diabetic current smoker had the worse survival after 20 months of follow-up, particularly between 24 months and 36 months.
- By eliminating these modifiable risk factors (MRFs), an estimated ~16% of the CS deaths could be avoided.

Introduction

Several MRFs, including diabetes, smoking, and BMI, are related to emerging pancreatic cancer. Epidemiological studies show that these MRFs also escalate mortality. Population-based studies assessing the impact of these MRFs on pancreatic cancer survival were limited. Studies which assessing these associations mainly controlled for sociodemographic factors only and showed inconsistent findings.

Objectives

1) To examine the impact of diabetes, smoking status, and BMI on pancreatic CS survival
2) To compare estimated survival rates among adult pancreatic cancer patients stratified by MRFs
3) To measure the population attributable risk (PAR) of these MRFs on survival outcome of pancreatic cancer patients

Table 1. Frequency, Hazard Ratios (HRs) and 95% CIs for Cancer-Specific Death

Variables	*n (%)*	*Crude HR (95% CI)*	*Adjusted HR (95% CI)*
Diabetes			
No	2,417 (65.2)	1	1
Yes	1,289 (34.8)	1.12 (1.04–1.21)	1.14 (1.05–1.23)
Smoking status			
Nonsmoker	1,422 (38.4)	1	1
Current smoker	857 (23.1)	1.24 (1.13–1.35)	1.39 (1.25–1.54)
Former smoker	1,015 (27.4)	1.14 (1.04–1.24)	1.14 (1.04–1.25)
Unknown	412 (11.1)	1.28 (1.14–1.44)	1.14 (1.00–1.31)
Body mass index (kg/m2)			
<18.5	172 (4.6)	1.20 (1.01–1.42)	1.00 (0.84–1.19)
18.5–<25	1,150 (31.0)	1	1
25–<30	998 (26.9)	0.84 (0.76–0.92)	0.93 (0.85–1.02)
30–<35	524 (14.1)	0.82 (0.73–0.92)	1.00 (0.90–1.13)
35–<40	211 (5.7)	0.79 (0.67–0.92)	1.02 (0.86–1.20)
≥40	136 (3.7)	1.01 (0.83–1.23)	1.46 (1.19–1.78)
Unknown	515 (14.0)	1.20 (1.07–1.34)	0.97 (0.86–1.10)

[a]Louisiana Tumor Registry/Epidemiology Program, School of Public Health, Louisiana State University Health Sciences Center. [b]Department of Public Health Sciences, College of Behavioral, Social and Health Sciences, Clemson University [c]Center for Outcomes and Health Services Research, Ochsner Health System.

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.

Methods and Materials

Data Source and Study Population

Data on pancreatic cancer patients diagnosed from 2011 to 2017 were queried from the Louisiana Tumor Registry (LTR). The eligibility criteria included pancreatic cancer patients aged 20 years and older with stage I–IV disease.

Modifiable Risk Factors

Diabetes mellitus data was retrieved from the patient's comorbid condition(s) and supplemented with statewide Hospital Inpatient Discharge Data (HIDD) 2010–2018 to obtain the complete information. Cigarette smoking, height and weight were abstracted directly from medical charts at the time of cancer diagnosis.

Sociodemographic and Clinical Variables

Race, age, marital status, insurance, census tract poverty, stage, grade, treatment, and Charson Comorbidity Index (CCI) score were included in the adjusted model.

Survival

Survival duration was defined as the time between the initial diagnosis date and the CS death date or end of follow-up, December 31, 2019, if alive. Patients who died from a non-CS cause were censored.

Statistical Analysis

The Cox regression model was used to examine the association between MRFs and CS survival. The stratified Cox regression model was used to estimate direct adjusted survival rates. The partial PAR was employed to measure the attributable risk of MRFs on CS survival.

Results

- Of the 3,706 eligible patients, 34.8% were diabetics, 23.1% were current smokers, and 50.4% had a BMI ≥ 25 kg/m^2.
- After adjusting for sociodemographic and clinical factors, diabetic patients had an increased CS death risk of 14% (95% CI, 1.05–1.23). The increased risk was 39% (95% CI, 1.25–1.54) for current smokers and 46% (95% CI,

Figure 1. Adjusted Hazard Ratios (HRs) and 95% CIs for Modifiable Risk Factors Stratified by American Joint Committee on Cancer (AJCC) Stages I–II (a) and Stages III–IV (b)

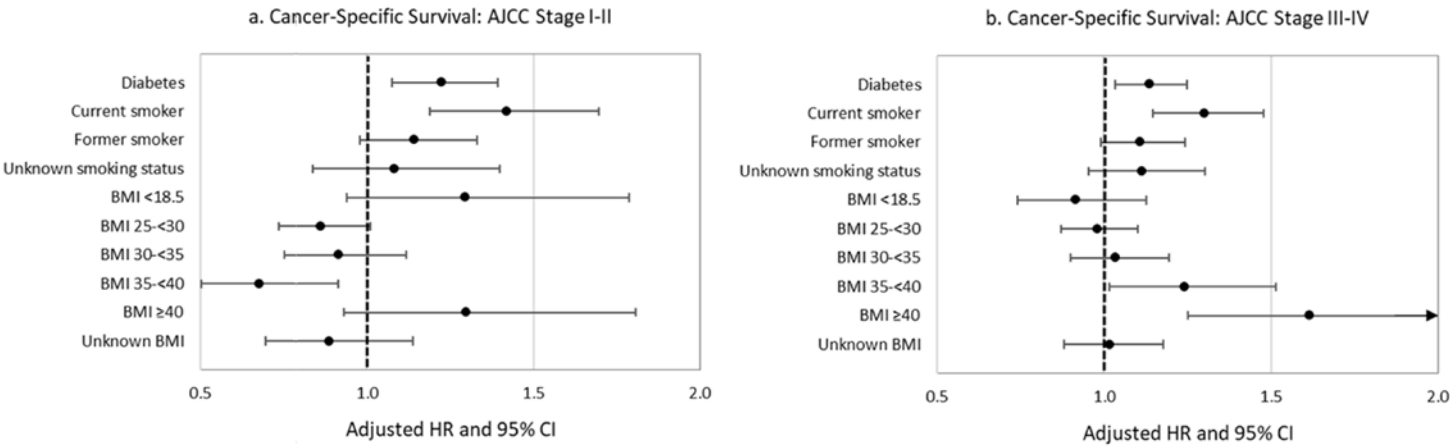
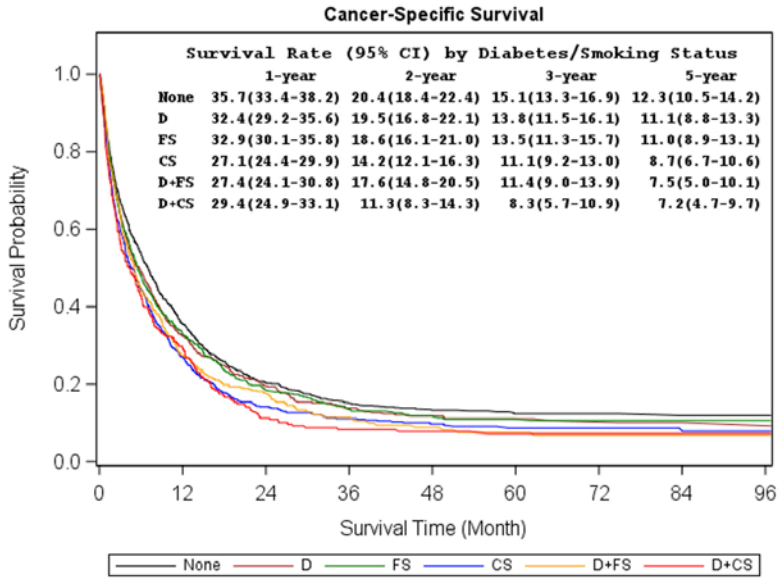


Figure 2. Adjusted Survival Curves for Pancreatic Cancer Patients by Diabetes and Smoking Status



D, diabetes only; CS, current smokers only; FS, former smokers only.

Table 2. Percent Population Attributable Risk (PAR) and 95% CIs of Diabetes, Smoking, and/or Body Mass Index (BMI) on Pancreatic Cancer-Specific Survival

Modifiable risk factors[a]	*Partial PAR% (95% CI)*[b]
Diabetes	4.5 (1.6–7.4)
Smoking	10.7 (5.6–15.8)
BMI	1.3 (0.5–2.1)
Diabetes, smoking	14.8 (6.8–22.6)
Diabetes, BMI	5.7 (1.4–10.1)
Smoking, BMI	11.9 (6.4–17.4)
Diabetes, smoking, BMI	15.9 (7.3–24.3)
Full PAR[c]	96.0 (88.0–98.7)

[a] Included cases with known smoking status and obesity (n = 3,001).

[b] One or more risk factors are considered eliminated, while others are allowed to remain unchanged.

[c] All pancreatic cancer patients who are exposed to risk factor(s) switch to the lowest risk category of all measured risk factors.

1.19–1.78) for patients with a BMI ≥ 40 when compared to their counterparts (Table 1).

- Diabetic patients and current smokers showed a significant increase in the risk of death which persisted after adjusting for covariates for both stage I–II and stage III–IV patients (Figure 1). However, a BMI ≥ 35 was observed to increase risk of mortality among stage III–IV patients only.
- Diabetic current smokers had significantly lower 2- and 3-year adjusted CS survival rates, 11.3% and 8.3% respectively (Figure 2).
- By eliminating MRFs, an estimated 15.9% (95% CI, 7.3%–24.3%) of the CS deaths could be avoided during the study period (Table 2).
- Among the 3 MRFs, smoking had the highest estimated partial PAR, 10.7% (95% CI, 5.6%–15.8%).

Conclusions

This study observed that diabetes and smoking contributed substantially to the reduction of pancreatic cancer survival after adjusting for sociodemographic and clinical factors; however, only BMI ≥ 35 kg/m^2 was observed to increase risk of mortality among stage III–IV patients. Advocacy and education on healthy lifestyle choices for the general population are imperative for cancer prevention and a favorable prognostic outcome.

Third Place Poster

Late-Stage Cervical Cancer Diagnosis in Young Adults in California following the Affordable Care Act

Julianne J. P. Cooley[a]; Frances B. Maguire[a]; Renata Abrahão[b,c]; Cyllene R. Morris[a]; Arti Parikh-Patel[a]; Theresa H. M. Keegan[a]

Background

- Young adults (YAs; ages 21–39 years), a historically underinsured population, may experience various barriers to healthcare, including lack of access to the HPV vaccine and Pap smear screening, which can prevent or detect premalignant lesions or cervical cancer at early stage (stage I).
- Following the Affordable Care Act (ACA), many YAs became eligible for insurance. However, YAs continue to be diagnosed with cervical cancer at later stages (II–IV).

Purpose

- To quantify changes in cervical cancer stage at diagnosis following the ACA and identify characteristics associated with later stage diagnosis.

Figure 1. Percent of Adenocarcinoma and Squamous Cell Carcinoma

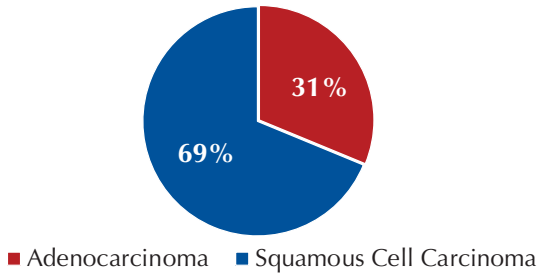


Figure 2. Percent of Late-Stage Diagnosis by Histology and Affordable Care Act (ACA) Implementation Period

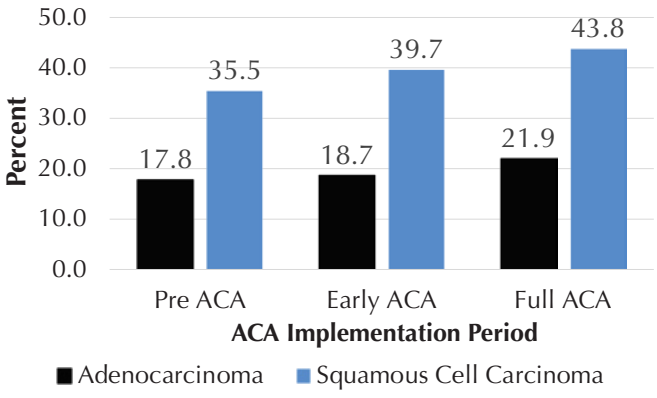
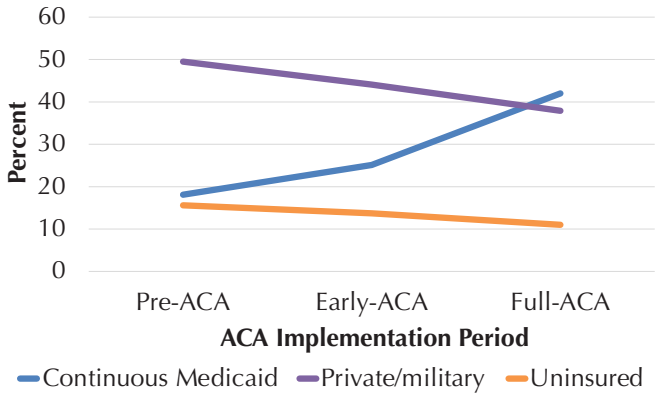


Figure 3. Percent of Health Insurance Type by Affordable Care Act (ACA) Implementation Period



Methods

- Using California Cancer Registry data linked to Medicaid enrollments, we identified YAs aged 21–39 years diagnosed with first primary squamous cell carcinoma (SCC) or adenocarcinoma (AC) cervical cancer pre-ACA (March 2005–September 2010), early-ACA (October 2010–December 2013), and post–full ACA implementation (January 2014–December 2017).
- Multivariable logistic regression was used to assess factors associated with later stage diagnosis in YAs diagnosed with AC or SCC. Results are presented as adjusted odds ratios (OR) and 95% CIs.

Results

- Of the 4,244 patients, 31% had AC and 69%, SCC (Figure 1).
- 32.7% of YAs were diagnosed at late stage. From pre-ACA to full-ACA, the percent of late-stage diagnoses increased by 6.5% (Figure 2).
- From pre- to full-ACA, continuous Medicaid coverage increased by 23%, whereas private insurance decreased by 11%, and Medicaid at diagnosis/uninsured decreased by 8% (Figure 3).
- YAs with Medicaid at diagnosis/uninsured, continuous Medicaid, and discontinuous Medicaid (vs private/military) were more likely to be diagnosed at a late stage for both AC and SCC histologies (Table 1).

[a]California Cancer Reporting and Epidemiologic Surveillance Program, University of California Davis Comprehensive Cancer Center, University of California Davis Health. [b]Center for Oncology Hematology Outcomes Research and Training (COHORT) and Division of Hematology and Oncology, University of California Davis School of Medicine. [c]Center for Healthcare Policy and Research, University of California Davis School of Medicine.

Address correspondence to Julianne J. P. Cooley. Email: jjpcooley@ucdavis.edu.

Table 1. Association between Demographic and Clinical Factors with Late-Stage (II–IV) Cervical Cancer Diagnosis for Young Adult Patients

Characteristic	*Adenocarcinoma OR (95% CI)*	*Squamous cell carcinoma OR (95% CI)*
Age group (vs 21–25 years)		
26–39 years	1.19 (0.58–2.44)	1.65 (1.19–2.30)
ACA implementation period (vs pre-ACA)		
Early ACA	0.83 (0.59–1.17)	1.08 (0.89–1.32)
Post ACA	1.05 (0.76–1.43)	1.39 (1.16–1.68)
Health insurance type (vs private/military)		
Continuous Medicaid	2.28 (1.56–3.33)	1.56 (1.27–1.92)
Discontinuous	2.6 (1.64–4.11)	2.26 (1.76–2.91)
Other public	2.4 (0.57–10.05)	0.95 (0.37–2.39)
Uninsured	2.89 (1.88–4.44)	3.23 (2.49–4.20)
Race/ethnicity (vs non-Hispanic White)		
Non-Hispanic Black	0.83 (0.35–1.97)	1.8 (1.28–2.53)
Hispanic	0.91 (0.66–1.27)	1.35 (1.11–1.63)
Asian/Pacific Islander	1.63 (1.08–2.45)	1.16 (0.87–1.54)
American Indian	1.3 (0.37–4.53)	0.59 (0.25–1.36)
Neighborhood socioeconomic status (vs high)		
Low	1.41 (0.96–2.07)	1.07 (0.85–1.35)
Medium	1.09 (0.77–1.54)	1.05 (0.83–1.31)
Rural residence (vs urban)		
Rural	0.545 (0.34–0.87)	0.91 (0.72–1.15)
Comorbidities (vs none)		
One comorbidity	1.17 (0.75–1.83)	1.31 (0.99–1.74)
More than 1 comorbidity	1.92 (0.64–5.82)	3.26 (1.91–5.57)
Marital Status (vs married)		
Not married	1.21 (0.91–1.61)	1.21 (1.02–1.43)
Care facility type (vs NCI-Designated)		
Non-NCI-designated	0.91 (0.67–1.24)	0.72 (0.60–0.87)

ACA, Affordable Care Act; NCI, National Cancer Institute; OR, odds ratio.

- In AC patients, Asian/Pacific Islanders (vs non-Hispanic Whites) were more likely to be diagnosed at later stage (Table 1).
- In SCC patients, older YAs, those of Black or Hispanic race/ethnicity (vs non-Hispanic White), patients with more than 1 comorbidity, and those diagnosed after the full ACA Expansion (vs pre-ACA) were more likely to be diagnosed at later stage (Table 1).

Conclusion

- Despite fewer YAs being uninsured and more continuously insured with Medicaid, the proportion of late-stage squamous cell carcinoma increased from pre- to post-ACA implementation.
- Our findings highlight the importance of access to the HPV vaccine and increased screening among underserved YAs in California.

Journal of Registry Management Continuing Education Quiz—FALL 2021

CANCER AMONG REFUGEES RESETTLED TO IDAHO DURING 2008–2019: PROOF OF CONCEPT STUDY

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

- Describe the general patterns of cancer burden among refugees in Idaho compared to the general Idaho population.
- Understand essential strengths and limitations of using the Centers for Disease Control and Prevention (CDC)'s Electronic Disease Notification (EDN) System for identifying cancer burden among refugees.

1. The cancer experience of refugees is similar to other socioeconomically disadvantaged populations in the United States.
   a) True
   b) False
   c) Unknown if true or false

2. The CDC's EDN System can be linked to cancer registry data not only to assess cancer burden among refugees, but also to supplement the country of birth variable (NAACCR Item #254).
   a) True
   b) False

3. The CDC's EDN System data are only available for Idaho, so this analysis cannot be replicated in any other state.
   a) True
   b) False

4. What years are refugee data available for the CDC's EDN System data?
   a) 2021 forward
   b) 2006 forward
   c) 2006–2016
   d) 2000–2016

5. Which data item was available in the CDC's EDN System but had a large proportion of nonspecific data?
   a) Social Security number
   b) International Medical Identifier
   c) Date of birth
   d) Name

6. The quality of the CDC's EDN System was high enough that manual review was not needed on potential matches.
   a) True
   b) False

7. Because no population denominators for refugees are available, cancer burden among refugees cannot be estimated.
   a) True
   b) False

8. Refugees in Idaho had a statistically significant, lower-than-expected number of malignant cancer diagnoses.
   a) True
   b) False

9. Refugees in Southeast Asia had higher rates of which cancer when compared to the general Idaho population?
   a) Prostate
   b) Breast
   c) Stomach
   d) Esophagus

10. Larger and more ethnically diverse states with greater proportions of refugees than Idaho will show similar cancer patterns among their refugee populations.
    a) True
    b) False
    c) Unknown if true or 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.

National Cancer Registrars Association
CALL FOR PAPERS

Danette A. Clark, BS, RMA, AAS, CTR | EDITOR-IN-CHIEF, *JRM*

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: Danette Clark, BS, RMA, AAS, CTR, Editor-in-Chief, *Journal of Registry Management,* (973) 971-5189, 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 the Danette Clark, BS, RMA, AAS, CTR, Editor-in-Chief at JRMEditor@ncra-usa.org or (973) 971-5189.

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.

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