Title

Codability of industry and occupation information from cancer registry records: Differences by patient demographics, casefinding source, payor, and cancer type

Journal

American Journal of Industrial Medicine, 61(6)

ISSN

0271-3586

Authors

Silver, Sharon R
Tsai, Rebecca J
Morris, Cyllene R
et al.

Publication Date

2018-06-01

DOI

10.1002/ajim.22840

Peer reviewed

Codability of Industry and Occupation Information from Cancer Registry Records: Differences by Patient Demographics, Casefinding Source, Payor, and Cancer Type

Silver SR[1], Tsai RJ[1], Morris CR[2], Boiano JM[1], Ju J[1], Scocozza MS[2], Calvert GM[1]

[1]National Institute for Occupational Safety and Health
Division of Surveillance, Hazard Evaluations, and Field Studies
[2]California Cancer Reporting and Epidemiologic Surveillance Program
Institute for Population Health Improvement
University of California Davis Health

Running title: Cancer Registry Industry/Occupation Codability

Corresponding author:
Sharon Silver
ssilver@cdc.gov

National Institute for Occupational Safety and Health
Division of Surveillance, Hazard Evaluations, and Field Studies
MS R-17
1150 Tusculum Ave.
Cincinnati, OH 45226
(513) 841-4313

Ms. Silver planned and implemented the analyses and wrote the manuscript. Ms. Tsai performed data management and provided extensive comments and revisions for the manuscript. Dr. Morris and Ms. Scocozza provided key information about the procedures of the California Cancer Registry and provided revisions to the manuscript. Mr. Boiano provided an industrial hygiene perspective during project development and provided comments and revisions for the manuscript. Ms. Ju led programming and data management for the project and provided review of the manuscript. Dr. Calvert planned and secured funding for the National Occupational Research Agenda project under which the data were received and provided extensive comments and revisions for the manuscript.

The NORA project was approved by the NIOSH Institutional Review Board (IRB) and was approved by the California Committee for the Protection of Human Subjects.

Acknowledgements (to be completed)

The authors would like to ~~that Pam Schumacher, [coders go here] for …, X and Y~~thank MaryBeth Freeman, Karla Armenti, and Marie Haring Sweeney for ~~their~~the valuable comments on the manuscript; Pam Schumacher for managing industry and occupation (I&O) coding for the project; Matthew Hirst for QA of I&O coding; Surprese Watts, Sue Burton, and Ashley Fite for I&O coding; and the California Cancer Registry personnel for providing the data used in these analyses.

NIOSH Disclaimer

The findings and conclusions presented in this article are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health. Mention of company names or products does not constitute endorsement by the National Institute for Occupational Safety and Health.

The authors have no conflicts of interest.

ABSTRACT

Introduction:

Industry and occupation (I&O) information collected by central cancer registries ~~are~~is useful for assessing associations among jobs and malignancies. However, systematic differences in I&O availablility can affect study findings.

Methods:

Codability by patient demographics, payor, identifying (casefinding) source, and cancer site was assessed using I&O text from first primary cancers diagnosed 2011-2012 and reported to the California Cancer Registry. I&O were coded to a U.S. Census code or classified as blank/inadequate/unknown, retired, or not working for pay.

Results:

Industry was codable for 37% of cases ~~but~~and blank/inadequate/unknown for 50%; another 9% had “retired” instead of usual industry. Cases initially reported by hospital sources, covered by preferred provider organizations, or with known occupational etiology (e.g. mesothelioma) were most likely ~~and~~to have codable industry, while cases initially reported by private pathology laboratories, Medicaid-covered cases, and malignancies frequently diagnosed in outpatient settings (e.g. melanoma) were least likely ~~to have codable industry~~. Results were similar for occupation.

Conclusions:

Recording usual I&O for retirees and finding additional sources for cases reported by entities without direct patient access would improve I&O codability and the validity of research findings.

Keywords:

cancer

registry

coding

occupation

industry

surveillance

INTRODUCTION

Population-based cancer registries are invaluable resources for research into multiple aspects of cancer and have often been used for assessing cross-sectional and longitudinal patterns of cancer incidence. Registry data have also facilitated preliminary assessment of associations among patient risk factors, including occupational exposures, and health outcomes. For example, industry and occupation data from the California Cancer Registry (CCR) have been used to evaluate the risk of cancer among firefighters,[1] the risk of leukemia subtypes[2] and lung cancer[3] among construction workers; and to assess differences in risk of acute myeloid leukemia by industry and occupation.[4] Given the importance of occupational exposures as risk factors for cancer, the 1992 Cancer Registries Amendment Act (1992 Act) included industrial and occupational (I&O) history among data items required to be collected, if available, for reported incident cases.[5]

However, if I&O information is not consistently available and of sufficient specificity to allow assignment to a standardized set of I&O codes, the validity, precision, and generalizability of the results of surveillance and epidemiologic analyses can be affected. The CCR requires that "every effort be made to record the I&O in which the patient works or worked," with the information ideally referring to the usual or longest held job,[6] a mandate more stringent than the 1992 Act, which requires only collection of I&O where available from the same record.[5] CCR sources of I&O information include admission and discharge summaries, face sheets, patient history, oncology consultation reports, and health and social history questionnaires the patient has completed. Still, I&O data are often missing from state registry records, or are recorded in such a way that precludes assignment of standardized codes that are comparable and therefore usable for analyses.[7]

A study of I&O availability (categorizing "retired" and "non-working" as available) in the New Hampshire State Cancer Registry found differences by demographic characteristics and by broad groupings of malignancy type and data source.[8] The current project extends this area of research by focusing on codability of I&O text to Census 2010 I&O codes by malignancy type and by demographic, source reporting, and payor characteristics.

The aims of the current analyses are to identify areas for improvement of I&O data collection and to identify types of analyses likely to be most affected by missing I&O data.

MATERIALS AND METHODS

CCR is California's statewide cancer surveillance system ~~and~~ has collected information on all cancers (except non-melanoma skin cancer and carcinoma in-situ of the cervix) diagnosed in California residents since 1988. ~~Data are collected~~ The CCR is housed within ~~managed by a~~the California Department of Public Health, which receives funding from the Centers for Disease Control and Prevention's National Program of Cancer Registries. A system of regional registries, ~~which are~~ also affiliated with the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) program, collects and submits cancer data to CCR. Information collected include patient demographics, diagnosis, tumor characteristics, types of treatment, and follow-up reports. The majority of cases are reported by hospital sources, but non-hospital sources, such as physicians and private pathology laboratories, provide cases as well. The registry collects information on primary and secondary payor, the entity financially responsible for covering the patient's costs, from each facility involved in the diagnosis or treatment ~~for~~of the tumor. Reports from each facility are retained (rather than overwritten); for this research, only the payor from the last facility treating the patient was analyzed. Inmates, children, and cases with the Veterans Benefits Administration as ~~payer~~payor were excluded.

The current study is part of a larger project funded by the ~~NIOSH's~~National Institute for Occupational Safety and Health's (NIOSH) National Occupational Research Agenda (NORA) to code I&O for cancer registry cases from six states, including California, and then evaluate the use of job-exposure matrices to assign exposures to these cases. The ~~NORA~~ project was approved by the NIOSH Institutional Review Board (IRB) and ~~was approved or exempted by the IRBs from participating states, including~~by the California Committee for the Protection of Human Subjects.

Industry and occupation data from 257,020 first primary cancers reported to the CCR in either 2011 or 2012 were processed by the NIOSH Industry and Occupation Computerized Coding System (NIOCCS) software.[9] NIOCCS processes industry and occupation text data using a two-step process: where possible, records are

automatically assigned standardized [United States (U.S.) Census] codes for industry and occupation; records that are not automatically coded due to I&O text ambiguity, incomplete I&O data, or other problems are presented to a human coder for computer-assisted coding.

For this analysis, I&O text from diagnosis years 2011-2012 CCR records was run through both steps of NIOCCS. The resulting I&O codes were then merged back into the CCR data. For each case, industry and occupation were eachseparately classified as codable (through autocoding or computer-assisted coding) if a specific U.S. Census code could be assigned (2007 industry codes, 2010 occupation codes). When I&O did not match a Census code, the case was assigned to one of three "uncodable" categories: 1) retired; 2) not working for pay; or 3) unknown/missing/blank/information inadequate to code. (referred to as "blank or uncodable" herein). "Not a paid worker" comprises students, volunteers, homemakers, and unemployed persons and is a legitimate Census code for occupation but not for industry, and because it. Because this designation provides no information about work-related exposures, occupations so coded were reassigned to "uncodable." Prevalences of these codability classifications were calculated for demographic (age, sex, race/ethnicity) and reporting (casefinding source, payor) categories.

For each case in the CCR, information, including I&O text, may come from one or multiple health entity sources. While evaluation of sources of I&O information would optimally account for all contributing sources, that information was not available for the current study. The CCR electronic record identifies the source that first identified the tumor (casefinding source). In addition, the CCR includes a variable that summarizes the overall best source for abstracting information about the tumor (henceforth called "best source"). The casefinding source variable separates sources into nine hospital and five non-hospital groupings, and offersoffering specificity within each of these. "Hospital" here indicates that the reporting sequence began in a hospital. Some hospital subgroupings, such as pathology department review, also include reporting from non-hospital entities (i.e. external pathology laboratories) that process and report results for specimens sent by hospitals. Two

hospital case identification sources, daily discharge review and disease index review, do not themselves ~~contain~~have I&O information but lead registrars to seek case information from records of other hospital ~~departmetns. The~~departments. One other casefinding source, quality control review, originates with the cancer registry. The seven categories for the "best source" variable are broader: hospital inpatient and managed care plans with comprehensive, unified medical records; radiation treatment centers and medical oncology centers; laboratory; private medical practitioner; nursing home, convalescent hospital, or hospice; autopsy only; and death certificate only.

The source of specific data fields, including I&O, is not available in the registry's summary record. Thus the I&O text in the CCR might come from the initial casefinding source, the overall best source, from another source, or from more than one source. In the absence of information specifying the source of I&O text, our primary analysis examined codability by initial casefinding source. Then, for records with non-hospital casefinding sources, we determined how codability differed when best source was also considered~~.~~, as the latter can reflect information sources beyond the initial casefinding source, was also considered. All analyses were conducted using SAS version 9.3 (SAS Institute).

RESULTS

Results for industry and occupation were similar (almost all categories for all metrics differed by <3%), so only data for industry are presented here. Slightly more than 37% of cases diagnosed in 2011-2012 were codable to a specific industry. Uncodable industry fell into several groupings. A relatively small group (4% of cases) was classified as "not a paid worker." "Retired" was the designation for 9% of cases. Most of the remaining 50% of cases included entries that stated were blank or ~~illegible or~~ had information that was illegible or too vague to code (responses such as "own business" or "family business" or responses that were somewhat more specific but could not be assigned 2010 Census codes).

Differences in codability by sex or race/ethnicity were relatively small compared to differences by age grouping (Table 1). Males were somewhat more likely (40%) to have codable industry than females (34%). This discrepancy may reflect the 5% higher prevalence among females of not being in the paid workforce~~; these non-paid individuals include students, volunteers, and homemakers.~~. Industry was codable for similar percentages of black and white non-Hispanic subjects. However, whites were 4% more likely to have industry listed as "retired", while for blacks, inadequate information was more likely to preclude coding (about 5%). Codability was slightly lower for Hispanics than for non-Hispanic blacks and whites.

Larger coding differences were seen by age category. In every age group, at least 45% of records had missing or inadequate industry information, and more than half of workers age 65 and above were in this category. Only about one-quarter of subjects younger than 25 had codable industry; this age group comprised by far the largest ~~number~~percentage of subjects classified "not a paid worker" ~~(> 25~~(27.4%). Industry was codable for nearly half of subjects aged 25-55, but the codable percentage declined in older age groups as the percentage classified as retired increased, with the decline accelerating in the 65-69 year old category (data not shown).

Table I. Codability of Industry by Sex, Race, and Age Category

Category	% of category (n)*	Industry Codability (% by category)			
		Retired	Not a paid worker	Blank or uncodable	Coded industry
*Sex***					
Male	47.6 (122223)	9.0	1.4	49.2	40.4
Female	52.4 (134757)	9.4	6.4	49.9	34.3
Race					
Non-Hispanic White	59.9 (153991)	10.1	3.4	47.4	39.1
Non-Hispanic Black	6.3 (16175)	6.0	3.2	52.2	38.5
Hispanic	19.3 (49696)	8.0	6.0	51.9	34.2
Asian/Pacific Islander	11.5 (29564)	9.6	4.8	48.5	37.0
American Indian	0.5 (1217)	6.2	5.3	52.3	36.2
Other/Unknown	2.5 (6377)	3.5	1.1	82.1	13.3
Age category (years)					
<25	1.0 (2435)	0.04	27.4	46.0	26.5
25-<35	2.8 (7175)	0.1	6.4	45.9	47.6
35-<45	6.5 (16763)	0.04	6.1	45.6	48.3
45-<55	16.0 (41061)	0.3	4.8	46.6	48.3
55-<65	25.7 (65998)	2.5	4.0	48.9	44.6
65-<75	25.4 (65331)	15.2	2.6	50.2	32.0
>=75	22.7 (58257)	20.4	3.1	53.6	22.9
Total	100 (257020)	9.2	4.0	49.6	37.2

*Percentages may not sum to 100 due to rounding.
**Other/transsexual/transgender not reported separately (because n<50) but included in total.

Most cancer cases (~~92~~91.7%) were initially reported to the registry by hospital casefinding sources (Table 2). ~~"Hospital" here indicates hospital origin of the reporting sequence. Some "hospital" subgroupings, such as pathology department review, include also reporting from non-hospital entities (i.e.~~ external pathology laboratories) that process and report results for specimens sent by hospitals. Hospital pathology department review was by far the largest source of casefinding, comprising 60% of all cases; industry was coded for 40% of records identified by this source. Of non-hospital sources, private pathology laboratories reported the largest percentage of all cases (~~5~~4.6%), but provided codable industry information for fewer than 15% of those cases. Some non-hospital sources comprised less than one percent of reports, but provided either very good (death certificate follow-up, >60% coded) or very poor (physician as casefinding source, <20% coded) industry information.

Table II. Codability of Industry by Case Identification (casefinding) Source

Casefinding Source*	% of all sources** (n)	Industry Codability (% by category) Retired	Not a paid worker	Blank or uncodable	Coded industry
Hospital Sources	*91.7 (235751)*	*9.4*	*4.1*	*48.2*	*38.4*
Reporting Hospital, NOS	19.9 (51070)	6.6	4.1	55.6	33.7
Hospital pathology department review	60.4 (155146)	9.2	4.1	46.5	40.2
Daily discharge review	0.25 (644)	2.2	8.4	61.5	28.0
Disease index review	8.5 (21883)	16.8	4.5	44.4	34.2
Radiation therapy department/center	1.7 (4426)	9.0	2.7	41.3	47.0
Outpatient chemotherapy	0.23 (590)	6.3	5.9	38.1	49.7
Diagnostic imaging/radiology	0.43 (1097)	14.0	3.8	36.1	46.0
Tumor board	0.17 (432)	10.9	5.8	47.4	35.9
Other hospital reporting source, including clinic	0.17 (446)	5.2	2.5	44.4	48.0
Non-hospital sources	*7.6 (19590)*	*7.1*	*2.7*	*67.1*	*23.1*
Physician report	0.28 (709)	12.8	1.6	66.2	19.5
Consultation-only or pathology-only ~~report~~Report	1.7 (4412)	7.2	1.0	66.7	25.1
Private pathology laboratory report	4.6 (11881)	7.0	1.3	77.2	14.5
Death certificate follow-back	0.83 (2140)	3.4	13.5	20.2	62.9
Other non-hospital reporting source	0.15 (390)	20.3	8.7	31.8	39.2
Quality control review	0.33 (842)	5.2	3.3	56.4	35.0
Missing/invalid source Information	0.33 (836)	11.6	5.1	42.0	41.3

NOS=not otherwise specified

* Sources with < 50 reported cases (laboratory reports, hospital rehabilitation service or clinic, nursing home initiated case, Coroner's Office records review, Managed care organization or insurance records, out-of-state case sharing) are not shown separately but are included in italicized category totals.

**Percentages may not sum to 100 due to rounding.

Industry was most frequently blank or uncodable when obtained from one of three non-hospital sources: private pathology laboratories ~~(77%),~~ consultation only/pathology only reports ~~(67%),~~ and physician reports ~~(66%).~~ However, cross-tabulation of these sources with the "best ~~information~~ source" variable showed marked differences between the three sources (Figure 1~~), cases with~~): when the "best ~~tumor~~ source ~~reflecting that~~

~~additional information~~" was ~~collected from hospitals~~a hospital or managed care ~~sources had markedly better~~source, industry codability was higher than ~~those with best information from~~when private ~~physician~~physicians or ~~laboratory sources.~~laboratories were designated as "best source."

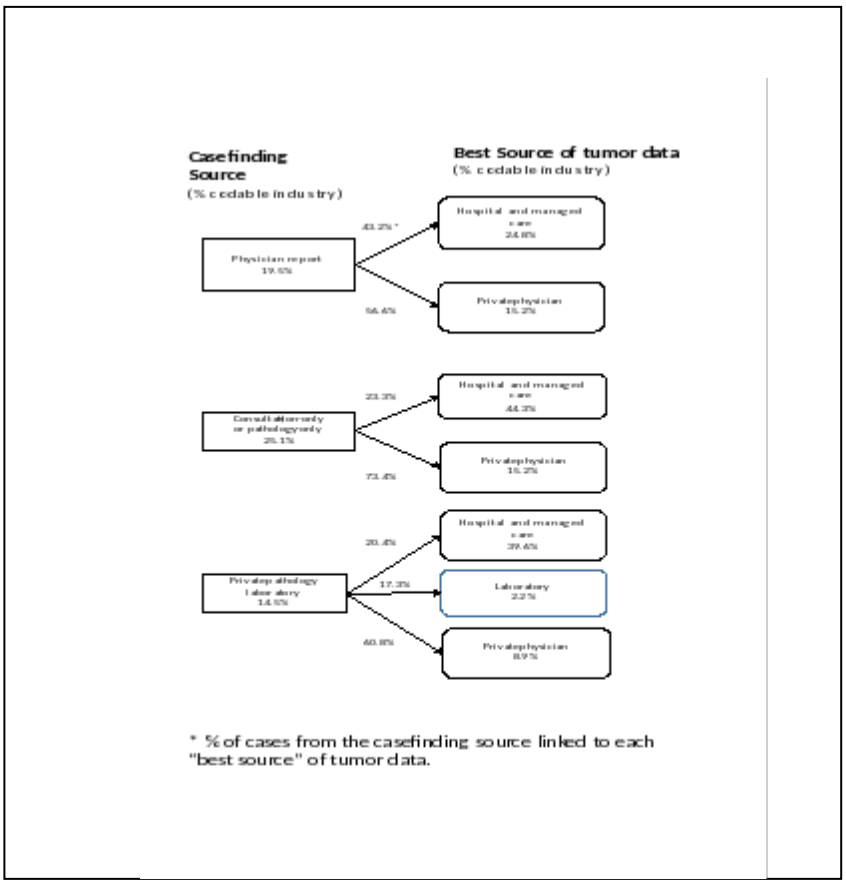

Figure 1: ~~How certain supplementary~~Supplementary sources can augment industry information for cases initially reported by entities with poor codability.

Industry codability differed by payor as well. HMOs, PPOs, and private insurance under managed care plans collectively were the payment source for 51.4% of cases (Table 3). Industry was reported for almost half of these cases, with cases covered by PPOs having the best codability (data not shown). In contrast, industry was codable for ~~less than one-fifth of cases~~only 17.6% involving dual-eligibility (patients eligible for Medicare and Medicaid). The dual-eligibility group had blank~~/inadequate~~ or uncodable industry information for ~~over half~~53.9% of cases

and ~~one-fifth~~had 20.7% of cases designated as "retired." ~~Prevalences~~When Medicaid was the provider or when the subject was uninsured, prevalence of codable industry ~~were~~was low (approximately 25% of reported cases) and ~~prevalences~~prevalence of missing/inadequate industry information high ~~when Medicaid was the provider or when the subject was uninsured.~~ .

Table III. Codability of Industry by Payor

Payor	% of all sources (n)*	Mean age at diagnosis	Industry Codability (% by category)			
			Retired	Not a paid worker	Blank or uncodable	Coded industry
Not insured /not-insured self-pay	2.0 (5033)	54.8	1.8	3.1	67.4	27.8
Insured: HMO, PPO, private insurance, or NOS	51.4 (132354)	58.6	5.1	3.4	44.6	46.9
Medicaid or covered by county	9.5 (24467)	53.5	2.8	7.7	63.8	25.8
Medicare: without supplemental Coverage or NOS	10.9 (27937)	73.4	17.6	4.0	52.3	25.5
Medicare: with supplemental coverage or via managed care	16.7 (42880)	74.4	18.7	3.0	46.5	31.8
Medicare with Medicaid eligibility	4.9 (12523)	71.8	20.7	7.8	53.9	17.6
Miscellaneous**	4.7 (12161)	63.6	4.3	3.6	67.5	24.6

NOS = not otherwise specified

*Percentages may not sum to 100 due to rounding.

**Unknown, military, or Indian Health Service

Industry codability for cancer sites comprising at least 1% of malignancies reported to the CCR is shown in Table 4; codability for all cancers reported to the CCR can be found in online appendix A. Three malignancies had industry coding rates above 50%: cancer of the tonsil (~~54~~53.9%, appendix A); pleural cancer (57.1%, data not shown); ~~pleural cancer (57%, n=14, data not shown);~~ and mesothelioma~~. Kaposi's~~ (54.0%). Kaposi sarcoma (~~data not shown~~appendix A) was least likely to have codable industry (27.1%) and most likely to have missing or inadequate industry (65.9%). Codability was below 33% (~~data not shown~~appendix A) for every cancer accessible on colonoscopy except anorectal cancer. For leukemia, codability varied by subtype, with acute lymphocytic

leukemia having the highest (45%, data not shown.1%, appendix A) and chronic lymphocytic leukemia the lowest (35%, data not shown.9%, appendix A).

Table 4. Industry Codability by Type of Malignancy, Descending Order by Codability

Malignancy*	% of all cases (n)**	Average age at diagnosis	Industry Codability (% by category)			
			Retired	Not a paid worker	Blank or uncodable	Coded industry
Oral Cavity and Pharynx	2.3 (5887)	61.2	6.5	2.9	46.6	44.0
Esophagus	0.8 (2064)	67.7	11.0	2.0	43.6	43.4
Thyroid and other endocrine	3.2 (8265)	49.2	4.0	7.1	45.8	43.2
Multiple Myeloma	1.3 (3390)	66.5	9.9	3.4	43.9	42.8
Breast	18.8 (48438)	60.0	7.5	6.7	44.4	41.4
Prostate	13.3 (34116)	66.0	8.2	0.5	50.9	40.3
Rectum	2.7 (7019)	61.6	9.2	3.6	47.2	40.0
Leukemia	2.2 (5648)	62.4	8.8	4.5	47.0	40.0
Non-Hodgkin Lymphoma	3.9 (10080)	63.6	9.2	3.7	47.8	39.4
Pancreas	2.5 (6331)	68.6	12.7	4.1	46.3	36.9
Lung, bronchus	8.6 (22131)	69.7	13.4	3.8	46.9	35.8
Kidney and renal pelvis	3.0 (7831)	62.2	8.9	3.5	52.3	35.3
Liver and intrahepatic bile duct	2.0 (5218)	64.1	8.8	4.4	51.8	35.0
Ovary	1.5 (3859)	60.2	9.2	5.9	49.9	34.9
Stomach	1.8 (4508)	65.8	13.2	4.1	48.1	34.7
Corpus Uteri, Uterus unspecified	3.4 (8864)	61.0	8.8	5.5	51.1	34.6
Cervix Uteri	1.0 (2561)	50.4	3.4	7.9	55.1	33.7
Colon and appendix	6.0 (15346)	67.1	12.7	3.6	53.0	30.8
Urinary bladder	3.5 (8918)	70.7	14.6	2.0	54.3	29.1
Melanoma of the skin/other non-epithelial skin cancers	7.1 (19104)	61.9	7.1	1.9	64.5	26.6

*Limited to sites comprising at least 1% of all malignancies reported to California Cancer Registry, 2011- 2012.

**Percentages may not sum to 100 due to rounding.

Differences in coding rates by malignancy type were not independent of distributions of reporting sources and/or payors (data not shown). For example, malignant melanoma and other non-epithelial skin cancers were six times as likely as all other cancers combined to be diagnosed by private pathology labs and ten times as likely as all cancers combined to be physician-initiated cases. Lung cancer, a common ~~outcome~~malignancy, had low codability, reflecting in part relatively large proportions of patients with “retired” recorded as industry and a higher than average percentage of cases paid by Medicare.

DISCUSSION

Results of this study, which found overall codability of I&O data to be less than 40%, point to the need for improvement in recording and collection of I&O data. In addition, the finding that I&O data availability varies by demographic, reporting, and outcome characteristics raises questions about the potential impact of missing I&O data on epidemiologic analyses of registry data. Research has shown that when data are not missing at random and missingness depends on both covariate values and disease status, biases may be introduced.[10] [11] The complete case approach, in which cases with incomplete data are excluded, has been shown to produce biased estimates with even 25% of data missing.[12] Multiple imputation, a common approach to missing data, is not feasible for filling in missing industry and occupation codes, with their large numbers of categories. No gold-standard population-based dataset that could be used to determine the actual distribution of cancer cases across industries and occupations exists. However, the current findings raise the possibility of selection bias on industry and occupation. If, for example, lung cancer incidence is 50% higher in industry A than in industry B, but industry information is 50% more likely to be missing for industry A (e.g. if most workers in industry A are uninsured and most workers in industry B have private/HMO/PPO insurance), lung cancer incidence would appear to be equal in the two groups, rather than double for industry A.

Increasing I&O codability, and therefore its utility for public health research, will require enhanced efforts in eliciting, recording, abstracting, and coding I&O. While this task appears daunting, recording usual I&O instead of "retired" would increase codability by nearly 10% in the CCR, and potentially to a greater extent in states like Texas, where "retired" has been reported to comprise 15% of cases from the Texas Cancer Registry.[13]

Codability varies by casefinding source. The finding in this assessment that private pathology labs and physicians reporting to the CCR have particularly low I&O codability echoes results from a study of New York State Cancer Registry data in which private physician offices and laboratories (not further specified) had the highest percentages of unknown values for race and Hispanic ethnicity, as well as tumor staging information.[14] The

impact of these deficiencies is particularly strong for melanoma, a malignancy for which diagnosis frequently involves submission of specimens by a dermatologist to an outside pathologist or pathology laboratory.[15] Electronic linkages between pathology labs and cancer registries are increasing,[16] but these pathology laboratories do not generally have any contact with patients, so linkage to the provider submitting the sample would likely be necessary in order to obtain I&O data. The administrative burden of seeking approval for such linkages and ensuring that they are executed is likely to be substantial, although increased use of electronic health records should facilitate these efforts. Targeted encouragement of physicians to report melanoma diagnoses has been suggested.[15] Such approaches (electronic linkages~~,~~ and educational efforts targeted to specific types of providers) are likely to be needed to encourage inclusion of I&O with reporting for melanoma and for other malignancies increasingly diagnosed in outpatient settings, such as colorectal malignancies found on colonoscopy. With this move to diagnosis in the outpatient setting, connecting to ambulatory care providers has become more important to ensure case completeness;[17] the same approach will be needed to enhance reporting of I&O by these sources. Electronic linkages will only improve I&O availability if these fields are consistently included and collected in electronic health records.

In the interim, hospitals might be a good starting point for efforts to improve I&O reporting, as they comprise the large majority of reports, are relatively centralized, and have some ~~units~~departments (those involved with the diagnosis and treatment of malignancies) reporting I&O at higher frequencies. ~~Information~~Some potential solutions to the lack of I&O availability include: information sharing within the hospital~~,~~ (including from departments such as registration, which are not linked directly to the medical records but collect job information for billing purposes); clearer instructions for eliciting I&O in registration systems and related paperwork~~,~~; and training for providers who collect the information to ensure that they consistently elicit and report usual specific I&O~~, could be helpful.~~ Specific barriers to reporting I&O in a 2005 study of Connecticut hospitals were lack of awareness of reporting requirements, lack of hospital reporting requirements, and insufficient time to report.[18] While some of these barriers may have shifted with the advent of electronic reporting, the lack of a standardized

requirement for inclusion of I&O information in the medical record is still perhaps the main challenge for the collection of meaningful I&O data.

Training of cancer registrars is also important. Cancer registrars may need to look at records from multiple sources to determine I&O, particularly when cases are initially identified by sources such as pathology labs that have no patient contact. Results of the current study show that such cases have markedly greater I&O codability when hospital records are also available and accessed for case information. A New Hampshire study showed a decrease in the number of records judged to have no I&O data from 74% to 14% after detailed records review followed by targeted registrar training in I/O collection.[8] The percentage of records judged to have complete, codable I&O data was only 48% even after review training (in part because 20% of records belonged to individuals not in the paid workforce), but this was nearly triple the level before these additional steps. However, the New Hampshire study noted that the location of I&O data within a medical record varied by type of facility and record system and that centralized training efforts might be more difficult to implement in larger states with more and varied healthcare facilities.

In the current study, records with death certificate follow-back as the casefinding source ~~with~~had the highest I&O codability ~~was death certificate follow-back.~~(though only 0.83% of cases had this source). However, the mixing of I&O data from death certificates with other sources of these data can be problematic for epidemiologic analysis. Death certificate data is, of necessity, obtained by proxy report (i.e., next-of-kin). A study comparing occupation from death certificates to occupation self-reported at midlife[19] found that while agreement for broad occupational categories was reasonable (67%), agreement for job titles was poor (32%). More importantly, death certificate data are only available for deceased cases, so their use would introduce additional potential for bias.

Pending efforts to improve collection of codable data, awareness of which study populations and outcomes are most affected by non-codability is important for planning epidemiologic studies. Non-codable data fell into

several categories, each impacting certain malignancies more strongly. Recording "retired" instead of usual industry generally had greater impact on malignancies occurring more frequently in older adults, such as bladder cancer. In contrast, young adults may be students or may not yet have established a "usual" occupation, leading to higher levels of categorization as "not a paid worker" for malignancies like testicular cancer that are more commonly diagnosed in younger adults. Low codability due to blank or uncodable I&O fields was associated with lack of insurance or with public coverage (Medicaid or county coverage); the high level of missing I&O among patients with Medicaid and county coverage could be associated with the reporting source (differences in information collection or reporting systems) and/or with characteristics of the covered case population (lower employment stability reducing the likelihood of a specific "usual" occupation and increasing likelihood of being unemployed at time of diagnosis; as well as disincentives to reporting employment). Designing studies to compare groups that are similar with respect to factors such as insurance coverage should decrease the potential for selection bias.

Several limitations pertain to this study. Collection of "usual" I&O is preferred to current I&O for research purposes (and a full work history would be better still), but the prevalence of "retired" in I&O fields suggests that directions on healthcare intake forms may not request "usual" I&O and that there may be suboptimal probing by providers, who have many competing demands. In addition, information collection forms may simply elicit "occupation" or "industry" or even "job" rather than specifying usual or longest-held I&O. Correlations between current and usual I&O have been found to be good for high-level I&O groupings, but the concordance decreased as more detailed I&O groupings were used.[20] Recommendations and/or incentives from healthcare accreditation organizations such as The Joint Commission, or by the American College of Surgeons Cancer Programs could increase awareness of the value of I&O data for cancer research and improve the regular collection of this information by healthcare providers.

Despite these limitations, cancer registry I&O data present the potential for meaningful assessment of associations between different types of work and malignancies. Findings of an examination of cancer outcomes among firefighters using CCR data[1] were generally consistent with those of a cohort study of firefighters.[21] While traditional cohort and case-control studies usually have access to more detailed work history information, primary advantages of analyzing registry data are the much lower cost and larger study populations.

Enhancing the utility of cancer registries for public health research by increasing I&O reporting will involve prioritizing data collection at the source, training of providers and registrars, and development of feedback loops to ensure continuous improvement. In the interim, consideration of these limitations while planning epidemiologic analyses of registry data is important to limit the potential for bias in the results.

References:

1. Tsai RJ, Luckhaupt SE, Schumacher P, Cress RD, Deapen DM, Calvert GM. Risk of cancer among firefighters in California, 1988-2007. Am J Ind Med. 2015;58(7):715-729.
2. Luckhaupt SE, Deapen D, Cress R, Schumacher P, Shen R, Calvert GM. Leukemia among male construction workers in California, 1988-2007. Leuk Lymphoma. 2012;53(11):2228-2236.
3. Calvert GM, Luckhaupt S, Lee SJ, et al. Lung cancer risk among construction workers in California, 1988-2007. Am J Ind Med. 2012;55(5):412-422.
4. Tsai RJ, Luckhaupt SE, Schumacher P, Cress RD, Deapen DM, Calvert GM. Acute myeloid leukemia risk by industry and occupation. Leuk Lymphoma. 2014;55(11):2584-2591.
5. Cancer Registries Amendment Act, 3372 (Public Law 102-515).
6. Kairon RS. California Cancer Reporting System Standards, Volume I: Abstracting and Coding Procedures. 2017; http://www.ccrcal.org/paqc_pubs/v1_2017_online_manual/index.htm. Accessed 5/26/2017, 2017.
7. Freeman MB, Pollack ~~L.~~LA, Rees JR, Johnson CJ, Rycroft RK, Rousseau DL, Hsieh MC. Capture and ~~Coding~~coding of ~~Occupation~~industry and ~~Industry Measures~~occupation measures: Findings from ~~10 Central~~eight National Program of Cancer Registries~~. NAACCR Annual Conference. St. Louis, MO2016.~~ states. Am J Ind Med. 2017;60(8):689-695.
8. Armenti KR, Celaya MO, Cherala S, Riddle B, Schumacher PK, Rees JR. Improving the quality of industry and occupation data at a central cancer registry. Am J Ind Med. 2010;53(10):995-1001.
9. Schmitz M, Forst L. Industry and Occupation in the Electronic Health Record: An Investigation of the National Institute for Occupational Safety and Health Industry and Occupation Computerized Coding System. JMIR Med Inform. 2016;4(1):e5.
10. Vach W, Blettner M. Missing Data in Epidemiologic Studies. Encyclopedia of Bioastatistics. Vol 5: John Wiley & Sons; 2005.
11. Heitjan DF. Incomplete data: what you don't know might hurt you. Cancer Epidemiol Biomarkers Prev. 2011;20(8):1567-1570.
12. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. BMC Med Res Methodol. 2010;10:112.
13. Weiss NS, Cooper SP, Socias C, Weiss RA, Chen VW. Coding of Central Cancer Registry Industry and Occupation Information: The Texas and Louisiana Experiences. J Registry Manag. 2015;42(3):103-110.
14. Kahn AR, Schymura MJ, Juster TM. Completeness of source-level data items based on type of source. NAACCR Annual Conference. St. Louis, MO2016.
15. Cockburn M, Swetter SM, Peng D, Keegan TH, Deapen D, Clarke CA. Melanoma underreporting: why does it happen, how big is the problem, and how do we fix it? J Am Acad Dermatol. 2008;59(6):1081-1085.
16. Gal TS, Durbin EB. Successes and challenges in population-based electronic pathology reporting. NAACCR Annual Conference. Quebec City, QC Canada2010.
17. Ayers CM, Celaya MO, Rees JR. Increasing non-hospital reporting: the NH experience. North American Association of Central Cancer Registries (NAACCR) Annual Conference. St. Louis, MO2016.
18. Polednak AP. Obtaining occupation as an indicator of patients' socioeconomic status in a popluation-based cancer registry. J Registry Manag. 2005;32(4):176-181.
19. Bidulescu A, Rose KM, Wolf SH, Rosamond WD. Occupation recorded on certificates of death compared with self-report: the Atherosclerosis Risk in Communities (ARIC) Study. BMC Public Health. 2007;7:229.
20. Luckhaupt SE, Cohen MA, Calvert GM. Concordance between current job and usual job in occupational and industry groupings: assessment of the 2010 national health interview survey. J Occup Environ Med. 2013;55(9):1074-1090.

21. Daniels RD, Kubale TL, Yiin JH, et al. Mortality and cancer incidence in a pooled cohort of US firefighters from San Francisco, Chicago and Philadelphia (1950-2009). Occup Environ Med. 2014;71(6):388-397.

Appendix A – Industry Codability by Detailed Malignancy (malignancies with at least 100* cases reported to California Cancer Registry 2011-2012)

	Industry Codability ~~(% by category)~~(%)	

Malignancy	% of all cases (n)*)**	Mean age at dx	Retired	Not paid worker or blank/ uncodable industry	Blank or uncodable	Coded industry
Oral Cavity and Pharynx						
–Lip	0.16 (412)	63.7	7.5 (31)	2.2 (9)	64.3 (265)	26.0 (107)
–Tongue	0.73 (1877)	61.6	6.3 (118)	2.4 (46).6	44.2 (829)	47.1 (884)
–Salivary gland	0.25 (649)	60.3	8.2 (53)	4.2 (27)50.0	45.8 (297)	41.9 (272)
–Floor of mouth	0.09 (230)	64.3	7.4 (17)	5.2 (12)56.1	50.9 (117)	36.5 (84)
–Gum and other mouth	0.25 (635)	65.5	9.0 (57)	2.4 (15)54.8	52.4 (333)	36.2 (230)
– Nasopharynx	0.18 (473)	54.4	4.7 (22)	4.0 (19)	4448.6 (211)	46.7 (221)
–Tonsil	0.42 (1077)	58.5	3.2 (34)	2.7 (29)42.9	40.2 (433)	53.9 (581)
– Oropharynx	0.08 (214)	63.4	9.3 (20)	1.9 (4)50.5	48.6 (104)	40.2 (86)
– Hypopharynx	0.10 (251)	65.2	10.4 (26)	3.6 (9)50.2	46.6 (117)	39.4 (99)
Other oral cavity and pharynx	0.03 (69)	62.0	4.4 (3)	1.4 (1)	58.0 (40)	36.2 (25)
Digestive System						
– Esophagus	0.80 (2064)	67.7	11.0 (227)	2.0 (41)	4345.6 (900)	43.4 (896)
–Stomach	1.75 (4508)	65.8	13.2 (595)	4.1 (183)52.2	48.1 (2167)	34.7 (1563)
–Small intestine	0.40 (1030)	63.9	10.2 (105)	3.4 (35)50.7	47.3 (487)	39.1 (403)
–Cecum	1.29 (3319)	70.4	15.3 (506)	3.1 (104)56.4	53.3 (1769)	28.3 (940)
–Appendix	0.17 (426)	57.7	7.5 (32)	451.7 (20)	47.0 (200)	40.8 (174)
– Ascending colon	1.11 (2855)	70.2	14.5 (413)	3.6 (104)56.9	53.3 (1523)	28.6 (815)
–Hepatic flexure	0.27 (699)	68.5	12.0 (84)	57.2.6 (18)	54.6 (382)	30.8 (215)
– Transverse	0.50 (1274)	69.2	12.9 (164)	357.6 (4)	54.0 (688)	29.5 (376)

colon						
–Splenic flexure	0.18 (472)	65.1	10.4 (49)	3.2 (15)57.0	53.8 (254)	32.6 (154)
– Descending colon	0.35 (910)	64.4	12.3 (112)	2.5 (23)54.9	52.4 (477)	32.8 (298)
–Sigmoid colon	1.81 (4651)	63.2	10.0 (466)	3.7 (172)57.0	53.3 (2478)	33.0 (1535)
–Large Intestine, NOS	0.29 (740)	69.7	16.2 (120)	6.5 (48)55.2	48.7 (360)	28.7 (212)
– Rectosigmoid junction	0.67 (1722)	62.9	10.7 (184)	3.7 (64)52.2	48.5 (835)	37.1 (639)
–Rectum	2.06 (5297)	61.6	8.7 (463)	50.3.6 (190)	46.7 (2476)	40.9 (2168)
–Anus, anal canal, & anorectum	0.72 (1838)	56.1	4.7 (86)	4.5 (83)55.8	51.3 (943)	39.5 (726)
–Liver	1.85 (4749)	63.4	8.4 (401)	4.4 (208)56.7	52.3 (2485)	34.9 (1655)
– Intrahepatic bile duct	0.18 (469)	66.0	12.8 (60)	4.5 (21)51.2	46.7 (219)	36.0 (169)
– Gallbladder	0.29 (749)	69.3	15.0 (112)	59.4.3 (32)	55.1 (413)	25.6 (192)
–Other biliary	0.4 (1143)	68.6	15.1 (172)	4.0 (46)52.7	48.7 (557)	32.3 (368)
–Pancreas	2.46 (6331)	68.6	12.7 (803)	50.4.1 (257)	46.3 (2934)	36.9 (2337)
– Retroperitoneum	0.07(187)	57.5	7.5 (14)	8.0 (15)	3947.6 (74)	44.9 (84)
– Peritoneum, omentum, & – mesentery	0.15 (379)	66.5	16.6 (63)	5.8 (22)50.1	44.3 (168)	33.3 (126)
–Other digestive organs	0.12 (310)	69.1	13.2 (41)	3.9 (12)58.1	54.2 (168)	28.7 (89)
Respiratory System						
–Nose, nasal	0.14 (367)	63.4	9.3 (34)	51.24.6 (17)	46.6 (171)	39.5 (145)

	cavity, & middle ear						
	–Larynx	0.52 (1343)	66.2	9.1 (122)	2.8 (37)49.3	46.5 (624)	41.7 (560)
	–Lung and bronchus	8.61 (22131)	69.7	13.4 (2965)	3.9 (853)50.8	46.9 (10384)	35.8 (7929)
–Pleura		0.01 (14)	68.1	0	7.1 (1)	35.7 (5)	57.1 (8)

Industry Codability (% by category)

Malignancy	% of all cases (n)*	Mean age at dx	Retired	Not paid worker	Blank or uncodable	Industry coded
–Trachea, mediastinum, & other respiratory organs	0.04 (93)	41.7	1.1 (1)	10.8 (10)	39.8 (37)	48.4 (45)
Bones and joints	0.15 (398)	49.1	5.0 (20)	10.0 (40)51.5	41.5 (165)	43.5 (173)
Soft tissue, including heart	0.67 (1727)	56.9	7.2 (125)	548.3 (91)	43.0 (744)	44.1 (767)
Skin excluding basal and squamous						
– Melanoma of the skin	7.05 (18117)	61.6	6.9 (1249)	1.8 (327)66.7	64.9 (11750)	26.4 (4791)
–Other non-epithelial skin	0.38 (987)	66.2	10.1 (100)	3.0 (30)61.1	58.1 (573)	28.8 (284)
Breast	18.85 (48438)	60.0	7.5 (3636)	6.7 (3232)51.1	44.4 (21511)	41.4 (20059)

Female Genital System

Industry codability (%)

Malignancy	% of all cases (n)*	Mean age at dx	Retired	Not paid worker or blank or uncodable industry	Industry coded (%)
–Cervix Uteri	1.0 (2561)	50.4	3.4 (86)	7.9 (203)63.0	55.1 (1410)	33.7 (862)
–Corpus Uteri	3.35 (8600)	61.0	8.9 (762)	56.5.4 (462)	51.0 (4389)	34.7 (2987)
–Uterus, NOS	0.10 (264)	61.8	7.2 (19)	962.1 (24)	53.0 (140)	30.7 (81)
–Ovary	1.50 (3859)	60.2	9.3 (357)	5.9 (229)55.8	49.9 (1927)	34.9 (1346)
–Vagina	0.12 (312)	61.0	10.3 (32)	3.8 (12)61.2	57.4 (179)	28.5 (89)
–Vulva	0.53	61.2	6.4 (86)	5.2	62.5 (844)	26.0 (351)

	(1351)			(70)67.7		
–Other female genital organs	0.13 (342)	58.6	12.9 (44)	7.0 (24)58.2	51.2 (175)	28.9 (99)
Male Genital System						
–Prostate	13.27 (34116)	62.9	8.2 (2801)	0.5 (174)51.4	50.9 (17375)	40.4 (13766)
–Testis	0.72 (1852)	66.0	0.4 (7)	5.8 (108)50.6	44.8 (830)	49.0 (907)
–Penis	0.12 (314)	34.5	8.9 (28)	2.2 (7)61.1	58.9 (185)	30.0 (94)
–Other male genital organs	0.03 (73)	62.7	12.3 (9)	0 (0)	56.2 (41)	31.5 (23)
Urinary System						
–Urinary bladder	3.47 (8918)	70.7	14.6 (1302)	2.0 (175)56.3	54.3 (4842)	29.1 (2559)
–Kidney & renal pelvis	3.05 (7831)	62.2	8.9 (698)	3.5 (271)55.8	52.3 (4097)	35.3 (2765)
–Ureter	0.10 (250)	73.1	16.0 (40)	3.2 (8)51.6	48.4 (121)	32.4 (81)
–Other urinary organs	0.05 (122)	71.7	15.6 (19)	454.1 (5)	50.0 (61)	30.3 (37)
Eye and orbit	0.19 (491)	63.8	7.1 (35)	2.0 (10)52.9	50.9 (250)	40.0 (196)
Brain and Other Nervous System						
–Brain	1.16 (2992)	57.3	6.6 (196)	5.1 (152)44.0	38.9 (1164)	49.5 (1480)
–Cranial nerves &other nervous system	0.07 (188)	53.9	5.9 (11)	48.4.8 (9)	43.6 (82)	45.7 (86)
Endocrine System						
–Thyroid	3.10 (7965)	49.0	3.9 (314)	7.1 (565)52.9	45.8 (3647)	43.2 (3439)
–Other endocrine including –thymus	0.12 (300)	55.8	4.7 (14)	6.7 (20)53.0	46.3 (139)	42.3 (127)
Lymphohaematopoietic						
–Hodgkin Lymphoma –Nodal- nodal	0.53 (1364)	42.3	3.1 (43)	9.9 (135)54.5	43.6 (594)	43.4 (592)
–Hodgkin Lymphoma –Extranodal	0.01 (13)	47.3	15.4 (2)	7.7 (1)	34.5 (5)	38.5 (5)
–Non-Hodgkin Lymphom	2.63 (6749)	63.7	9.7 (652)	3.6 (244)50.7	47.1 (3179)	39.6 (2674)

a —— —— Nodal					
Non-Hodgkin Lymphoma - Extranodal	1.30 (3331)	63.5	8.2	53.0	38.8
Myeloma	1.32 (3390)	66.5	9.8	47.3	42.8
Acute Lymphocytic Leukemia	0.22 (559)	45.0	2.9	52.0	45.1
Chronic Lymphocytic Leukemia	0.78 (2015)	69.0	10.7	53.4	35.9
Other Lymphocytic Leukemia	0.08 (198)	61.0	4.6	48.5	47.0
Acute Myelogenous Leukemia	0.66 (1685)	61.3	9.6	49.3	41.1
Chronic Myelogenous Leukemia	0.30 (782)	58.8	6.5	51.3	42.2
Aleukemic, subleukemic, & NOS	0.07 (185)	66.6	13.5	53.5	33.0
Mesothelioma	0.17 (435)	71.2	9.4	37.0	54.0
Kaposi Sarcoma	0.13 (340)	47.7	3.5	69.4	27.1
Miscellaneous	2.89 (7425)	70.5	13.3	57.5	29.0

			Industry Codability (% by category)			
Malignancy	% of all cases (n)*	Mean age at dx	Retired	Not paid worker	Blank or uncodable	Coded Industry
Non-Hodgkin Lymphoma — Extranodal	1.30 (3331)	63.5	8.2 (274)	3.8 (125)	49.2 (1639)	38.8 (1293)
Myeloma	1.32 (3390)	66.5	9.8 (335)	3.4 (115)	43.9 (1488)	42.8 (1452)
Acute Lymphocytic Leukemia	0.22 (559)	45.0	2.9 (16)	11.6 (65)	40.4 (226)	45.1 (252)
Chronic Lymphocytic Leukemia	0.78 (2015)	69.0	10.7 (216)	2.6 (52)	50.8 (1023)	35.9 (724)
Other Lymphocytic Leukemia	0.08 (198)	61.0	4.6 (9)	1.5 (3)	47.0 (93)	47.0 (93)
Acute Myelogenous Leukemia	0.66 (1685)	61.3	9.6 (162)	4.3 (73)	45.0 (758)	41.1 (692)
Chronic Myelogenous Leukemia	0.30 (782)	58.8	6.5 (51)	5.0 (39)	46.3 (362)	42.2 (330)
Other myeloid/monocytic leukemia	0.02 (63)	64.8	7.9 (5)	9.5 (6)	47.6 (30)	34.9 (22)
Acute Monocytic leukemia	0.03 (83)	63.6	7.2 (6)	2.4 (2)	41.0 (34)	49.4 (41)
Other acute leukemia	0.03 (78)	67.9	9.0 (7)	9.0 (7)	51.3 (40)	30.8 (24)
Aleukemic, subleukemic, & NOS	0.07 (185)	66.6	13.5 (25)	4.9 (9)	48.6 (90)	33.0 (61)
Mesothelioma	0.17 (435)	71.2	9.4 (41)	3.0 (13)	34.0 (146)	54.0 (235)
Kaposi Sarcoma	0.13 (340)	47.7	3.5 (12)	3.5 (12)	65.9 (224)	27.1 (92)
Miscellaneous	2.89 (7425)	70.5	13.3 (985)	3.8 (284)	53.9 (4001)	29.0 (2155)

NOS = not otherwise specified; dx = diagnosis

*The following reportable malignancies had fewer than 100 cases in CCR data for 2011-2012 and are not shown: Cancers of other oral cavity and pharynx sites; pleural cancer; cancers of other male genital organs; Hodgkin Lymphoma extranodal; other myeloid/monocytic leukemias; acute monocytic leukemia; other acute leukemia.

**Percentages may not sum to 100 due to rounding.