AMERICAN JOURNAL OF INDUSTRIAL MEDICINE 57:153–162 (2014)

Utility of Death Certificate Data in Predicting Cancer Incidence

Ronald L. Bedford, MS,[1] Spencer G. Lourens, BS,[2] Charles F. Lynch, MD, PhD,[3] Brian J. Smith, PhD,[2*] and R. William Field, PhD, MS[1,3]

Background *Studies often rely on death certificates to identify cancer occurrence. This research assessed the death certificate's ability to reflect cancer incidence and factors that influence agreement with cancer registry data.*
Methods *This study compared death certificates to cancer incidence data for an occupational cohort of 1,795 deceased workers who were registered by the Iowa Cancer Registry (ICR) between 1973 and 2005. Logistic regression models examined the effects of factors such as survival time, age at diagnosis, and gender on the odds of agreement between death certificate and incidence data.*
Results *Death certificates under-reported cancer incidence by 10–100%, depending on site. A 1-year increase in survival decreased the odds of agreement between death certificate and ICR data by 18%. Younger and female workers had increased odds of agreement.*
Conclusions *Death certificates can be useful predictors of cancer incidence, particularly for diseases with shorter survival and among subjects diagnosed earlier in life.* Am. J. Ind. Med. 57:153–162, 2014. *© 2013 Wiley Periodicals, Inc.*

KEY WORDS: cancer; death certificate; cancer incidence; underlying cause of death; all conditions coded

INTRODUCTION

Retrospective occupational epidemiology studies often rely on death certificate information to identify cancer occurrence. Use of cancer incidence data, however, provide greater power and sensitivity to detect associations between environmental/lifestyle exposures and cancer-related outcomes than mortality data alone. In the United States, the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) Program has collected information on cancer occurrence and mortality since as early as 1973, covering 28% of the population currently [NCI, 2008]. Established by Congress in 1992, the Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries (NPCR) provides resources and legislative support to central cancer registries representing 96% of the U.S. population. Today combined the two programs collect comprehensive cancer data on the entire U.-S. population. Unfortunately, many of these population-based registries have only recently been established or fully supported within the past two decades [NPCR, 2008].

Studies examining the agreement between death certificate and cancer registry data are limited in number. Of the few studies that have been performed, three [Percy et al., 1981, 1990; Demers et al., 1992] relied solely on the underlying cause of death (UCOD), neglecting the contribution of other cause of death (COD) information on the death certificate to assess the agreement between death certificate and cancer

[1]Department of Occupational and Environmental Health, College of Public Health, The University of Iowa, Iowa City, Iowa
[2]Department of Biostatistics, College of Public Health, The University of Iowa, Iowa City, Iowa
[3]Department of Epidemiology, College of Public Health, The University of Iowa, Iowa City, Iowa
Disclosure Statement: The authors report no conflicts of interests.
*Correspondence to: R. William Field, PhD, MS, University of Iowa, S327 CPHB, Iowa City, IA 52240. E-mail: bill-field@uiowa.edu

Accepted 20 July 2013
DOI 10.1002/ajim.22242. Published online 22 August 2013 in Wiley Online Library (wileyonlinelibrary.com).

registry data. Two other research teams [Sigurdson et al., 2003; Freedman et al., 2006] included cancers identified exclusively on death certificates to examine the agreement between death certificate and cancer registry data and were limited to reporting only overall cancer under-reporting figures based on death certificates.

The purpose of this report is to assess the ability of death certificates to accurately reflect cancer incidence within a population of decedents. Better understanding of how well cancer incidence can be ascertained from death certificates was a priority need in the parent study from which this report arose, the Department of Defense-funded Iowa Army Ammunition Plant (IAAAP) Munitions Workers Study. Production started at the IAAAP in 1941. Mortality analyses were performed for workers employed between 1946 and 2007. Deaths among the cohort occurred as early as 1952, over 20 years before the ICR, one of the original SEER registries, began collecting data in 1973 and 40 years before the NPCR was established in 1992. "Before NPCR was established, 10 states had no registry, and most states with registries lacked the resources and legislative support they needed to gather complete data" [NPCR, 2008]. We expect that researchers relying exclusively on death certificate data to identify cancer occurrence among retrospective cohort workers in areas only recently covered by population-based cancer registries as well as those examining time frames that predate the establishment of central cancer registries will gain a greater appreciation for the total burden of cancer in their cohorts.

Most studies examining cancer occurrence among cohorts not covered by central cancer registries are limited to death certificate data for cancer case ascertainment, commonly the UCOD only. This report presents a novel approach to quantify the extent of under-reporting of cancer occurrence by death certificates among cohorts that predate population-based cancer registries. Use of the UCOD will be compared to all conditions coded on the death certificate by evaluating the extent of agreement with ICR incidence data.

MATERIALS AND METHODS

Research Design

The study utilized existing data from a retrospective cohort epidemiology study that examined associations between occupational exposures and adverse health effects, including cancer incidence and mortality. The cohort comprised 36,953 workers employed at the Iowa Army Ammunition Plant (IAAAP) at Middletown, Iowa between January 1, 1941 and December 31, 2005. The research project received human subject approval from the University of Iowa's Institutional Review Board. Cohort identification and validation procedures have been detailed elsewhere [Bedford, 2009].

The Iowa Cancer Registry (ICR), a SEER registry since inception of the NCI's SEER Program in 1973, maintains comprehensive information about cancer incidence among Iowa residents. The ICR has consistently achieved 95% or greater completeness of its surveillance efforts in the state and reported almost 99% follow-up for 2012, the most recent reporting year [West et al., 2012]. The ICR also has access to death certificate information on all deaths certified in Iowa dating back to 1973, in the form of computerized mortality data provided by the Iowa Department of Public Health (IDPH) that includes information on COD.

Subject Inclusion Criteria

Study subjects met the following inclusion criteria: (1) employment at the IAAAP for at least one day between January 1, 1941 and December 31, 2005; (2) registered by the ICR with a first primary malignant neoplasm between January 1, 1973 and December 31, 2005; and (3) confirmation of death by the IDPH between January 1, 1973 and December 31, 2005. Both male and female former workers of any age were included. Cohort members were linked to ICR cancer incidence records using the CDC's Link Plus probabilistic record linkage software [NPCR, 2009]. Variables used for matching included name, sex, race, birth date, date of death, state of death, and zip code of last residence, if available. Records that did not meet Link Plus's default matching criteria were manually reviewed to determine linkages.

Subject Exclusion Criteria

Workers with non-invasive, including in situ, tumors were excluded from the study as these neoplasms are not expected to be fatal. Since cancer victims with multiple primary tumors are generally considered to have different mortality experiences than those with single primary neoplasms, subjects with multiple primary cancers were excluded. Workers with cancer types not included among the 18 sites readily discernible at the three-digit International Classification of Diseases (ICD) level were also excluded. Finally, workers whose cancers were identified by death certificate only were excluded from analysis due to a lack of verifiable diagnostic data from any source, including the ICR, to confirm the cancer.

Determination of UCOD and All Conditions Coded

The original purpose of determining and documenting a single UCOD on death certificates was to enable statistical analyses of COD data [Hetzel, 2008; World Health Organization, 2008; Moriyama et al., 2011]. As statistical analyses have become more sophisticated, analysis of multiple causes of death have become possible. Despite this capability, many studies still rely on the UCOD only for cancer incidence investigation.

Death certificate-reported COD information included: (1) the immediate COD, defined as the final disease or condition resulting in death (e.g., an organ system failure, neoplasm, injury, etc.) and if applicable, (2) a sequential list of conditions leading to the immediate cause that culminates with the UCOD, defined as the disease or injury that initiated the chain of events resulting in death; and (3) a list of other significant conditions contributing to death, but not resulting in the UCOD [National Center for Health Statistics, 2003]. Along with the UCOD, these contributing causes and significant other conditions are referred to as all conditions coded by the IDPH and the National Center for Health Statistics (NCHS). All medical conditions were coded by nosologists at the IDPH to the ICD revision in use at the time of death: ICD-8 for deaths before 1979, ICD-9 for deaths occurring from 1979 to 1998, and ICD-10 for deaths since 1999.

Identification of Incident Cancer Cases

Tumors recorded by the ICR were coded to the ICD for Oncology (ICD-0) revision in use at the time of diagnosis and subsequently updated to ICD-O-3, the most current revision [World Health Organization, 2000]. These diagnoses were compared to the UCOD and all conditions coded in the ICD mortality code revision in use at the time of the worker's death. Corresponding COD codes and ICD-O-3 site codes are presented in Appendix Table SI.

Statistical Analyses

To evaluate the extent of agreement between death certificates and registry data, the UCOD and all conditions coded were compared to the site of the first primary malignant neoplasm identified by the ICR for the 18 eligible cancer sites. ICR data served as the reference for these analyses, under the assumption that the ICR, which has access to medical records and pathology reports, had accurate information. The UCOD and ICR cancer incidence data were considered to agree if the UCOD was equivalent to the ICD-O-3 site code recorded by the ICR. The one-to-one correspondence between an UCOD and the site of a first primary malignant neoplasm made a subject-specific approach possible throughout the UCOD analyses.

We anticipated that the use of all conditions coded may increase the likelihood of identifying incident cancers. However, the potential to have multiple cancers listed on an individual death certificate required different analytic methods than were used to evaluate agreement between the UCOD and ICR incidence data. All conditions coded was considered to agree with the ICR incidence data if a COD code equivalent to the ICR site code was found anywhere among all conditions coded.

Percent under-reported and predictive value-positive (PV+) were chosen as measures of agreement between the UCOD and the cancer incidence data reported by the ICR. Since all conditions coded had the potential to identify the first malignant primary correctly as well as incorrectly for the same subject, calculation of the PV+ would have been invalid. For this reason, only correct matches between all conditions coded and the ICR incidence data were included to evaluate agreement between the sources as well as for the regression analysis. This subject-specific approach eliminated over-reporting by all conditions coded. Consequently, percent under-reported alone was used to evaluate the extent of agreement between all conditions coded and the cancer incidence data reported by the ICR.

Percent under-reported was defined as the probability that a particular neoplasm was not listed as the UCOD or among all conditions coded on the subject's death certificate given that the cancer was recorded by the ICR as the subject's first primary malignancy. To calculate this proportion, for each particular cancer, the number of death certificates that failed to match the ICR diagnosis determined the numerator with the count of study subjects identified by the ICR with that neoplasm in the denominator.

PV+ was the probability of occurrence of a particular cancer, given its identification as the UCOD. In this study, PV+ was the count of workers with a specific neoplasm reported as their UCOD, which agreed with ICR records, in the numerator and the total number of workers with that cancer listed as their UCOD in the denominator.

Percent under-reported or the false negative rate is also the complement of sensitivity. We noted in a study employing a similar cross-classification [Field et al., 2004], that sensitivity and PV+ provide the clearest indication of agreement in the comparison of cancer diagnoses because they measure the probability of correctly reflecting the occurrence of a site-specific neoplasm. Percent under-reported quantifies the extent to which mortality data fail to ascertain cancer incidence more directly than sensitivity.

Twelve cancers had totals of fewer than 20 workers in the cells representing the discordant pairs in the 2×2 tables constructed to evaluate the UCOD or all conditions coded relative to the ICR incidence data. For this reason and since computing software is available that can easily handle the calculations, the exact binomial test was used to test for differential misclassification for all neoplasms. Since all conditions coded could include more than one COD, a disease-specific approach was utilized to perform this testing. All conditions coded were searched for site codes corresponding to any of the 18 diseases studied. Results were compared to the ICR site code to evaluate under-reporting.

Multivariate logistic regression models were constructed to examine the effects of survival time, age at diagnosis, ICD revision in use at the time of death, diagnosis era, race, and sex on the odds of agreement between cancer occurrences identified as the UCOD and among all conditions coded on death certificates versus cancer incidence recorded by the ICR for the study population.

Survival time was calculated as a continuous variable by subtracting the worker's date of cancer diagnosis from the date of death recorded by the ICR. ICR records included the day, month, and year of death from the IDPH mortality database for all subjects, but only the month and year of diagnosis. All study subjects were assigned mid-range values of 15 for day of diagnosis. A mid-range value of six was assigned for workers with a missing value for month of diagnosis.

In the regression models, age at diagnosis was a discrete variable with a value equal to the worker's age at the time of cancer diagnosis. However, quartiles of the integral values (22–60, 61–68, 69–74, 75 or more years) produced better-fitting, more parsimonious models. Three different ICD revisions had been used to code CODs during the 33-year study time frame. To evaluate the possible effects of changes in medical diagnosis and treatment over time, the workers were divided into tertiles based on the worker's year of diagnosis.

Statistical analysis was performed using the SAS 9.1 statistical software (SAS Institute, Inc., Cary, NC). All statistical tests were two-sided and assessed for significance at the 5% level.

TABLE I. Demographic and Descriptive Information for the IAAAP Cohort Identified by the ICR With a First Primary Incident Malignant Neoplasm Between 1973 and 2005 and Matched to Mortality Information Provided by IDPH for the Same Time Period

Characteristic	n	%
Sex		
Male	1,260	70.2
Female	535	29.8
Race		
White	1,753	97.7
Black	40	2.2
American Indian, Aleutian, Eskimo	1	0.1
Invalid code	1	0.1
ICD revision		
ICD-8 (1973–1978)	268	14.9
ICD-9 (1979–1998)	1,222	68.1
ICD-10 (1999–2005)	305	17.0
Diagnosis era		
1973–1983	518	28.9
1984–1994	731	40.7
1995–2005	546	30.4

Characteristic	n	Mean (SD)	Minimum	Q1	Median	Q3	Maximum
Age at diagnosis	1,795	66.8 (11.2)	22	60	68	74	95
Age at death	1,795	69.5 (11.3)	32	62	70	77	100
Survival (years)	1,795	2.6 (4.2)	−0.3[a]	0.3	1.0	3.0	29.8

IAAAP, Iowa Army Ammunition Plant; ICR, Iowa Cancer Registry; IDPH, Iowa Department of Public Health.
[a]Sixteen workers had missing month of diagnosis; mid-range value of year (6 months) was assigned.

RESULTS

A subgroup of 2,684 former IAAAP workers had both a cancer identified by the ICR and a death certified by the IDPH between January 1, 1973 and December 31, 2005. A total of 889 subjects were excluded for at least one exclusion criterion with 145 workers meeting multiple criteria. Eighty-nine workers had non-invasive (including in situ) tumors and 411 eligible workers were registered with multiple primary cancers. An additional 482 workers had tumors which were not distinguishable at the three-digit ICD level. Finally, 52 workers had cancers identified only by death certificate that could not be verified by the ICR. A total of 1,795 former workers were included in the study.

Subjects were predominantly male (70%) and overwhelmingly white (98%) (Table I). Among this cohort, mean and median age at cancer diagnosis was 67 and 68 years, respectively, with a range from 22 to 95 years. The mean and median ages at death were 69 and 70 years, respectively, with a minimum of 32 and maximum of 100 years. Sixteen workers had missing values for month of diagnosis. These workers were assigned mid-range values of month 6. This, along with the assigned value for day of diagnosis, resulted in negative survival time values for approximately 1% of the cohort. Median survival time was slightly less than 1 year and the maximum was nearly 30 years.

Lung cancer was the most frequently recorded neoplasm with 739 incident cases which comprised 41% of the ICR registrations among the study cohort. Lung cancer was listed as the UCOD on 635 death certificates, with 628 (35%) matching ICR records (Table II). Cancer of the lung and bronchus was identified among all conditions coded on 665 (37.1%) death certificates (Table III). The rarest neoplasm among IAAAP workers was testicular cancer, which was recorded five times by the ICR, but was never listed as a UCOD or among all conditions coded. Overall, the UCOD failed to detect 566 (31.5%) incident cancers recorded by the

TABLE II. Cross-Classification of UCOD From Death Certificate and Site of First Incident Primary Malignant Neoplasm Reported by the ICR

Cancer identified as UCOD	First malignant primary tumor site identified by Iowa Cancer Registry (frequency (%))																	
	Esophagus	**Stomach**	**Colon and rectum**	**Liver and IBD**	**Pancreas**	**Larynx**	**Lung and bronchus**	**Skin melanoma and other MN**	**Breast**	**Cervix uteri**	**Corpus and uterus, NOS**	**Ovary and other uterine adnexa**	**Prostate**	**Testis**	**Urinary bladder**	**Kidney and renal pelvis**	**Brain and ONS**	**Thyroid**	**Total**
Esophagus	30 (1.7)	4 (0.2)																	34 (1.9)
Stomach	1 (0.1)	16 (0.9)	1 (0.1)																18 (1.0)
Colon and rectum		1 (0.1)	154 (8.6)		2 (0.1)		2 (0.1)												159 (8.9)
Liver and IBD			1 (0.1)	12 (0.7)	1 (0.1)		1 (0.1)												15 (0.8)
Pancreas				1 (0.1)	74 (4.1)		2 (0.1)												77 (4.3)
Larynx						7 (0.4)													7 (0.4)
Lung and bronchus	1 (0.1)	1 (0.1)	2 (0.1)			1 (0.1)	628 (35.0)						2 (0.1)						635 (35.4)
Skin melanoma and other MN						1 (0.1)		13 (0.7)											14 (0.8)
Breast									70 (3.9)										70 (3.9)
Cervix uteri										16 (0.9)									16 (0.9)
Corpus and uterus, NOS										4 (0.2)	9 (0.5)								13 (0.7)
Ovary and other uterine adnexa												21 (1.2)							21 (1.2)
Prostate													82 (4.6)						82 (4.6)
Testis														0 (0)					0
Urinary bladder	1 (0.1)		1 (0.1)												25 (1.4)	2 (0.1)			29 (1.6)
Kidney and renal pelvis							1 (0.1)						1 (0.1)		2 (0.1)	33 (1.8)			37 (2.1)
Brain and ONS							3 (0.2)		1 (0.1)								37 (2.1)		41 (2.2)
Thyroid																		2 (0.1)	2 (0.1)
ICR site code not listed as UCOD	7 (0.4)	4 (0.2)	106 (5.9)	2 (0.1)	10 (0.6)	17 (1.0)	102 (5.7)	14 (0.8)	41 (2.3)	6 (0.3)	13 (0.7)	5 (0.3)	126 (7.0)	5 (0.3)	33 (1.8)	22 (1.2)	5 (0.3)	7 (0.4)	525 (29.3)
Total underreported	10 (0.6)	10 (0.6)	111 (6.2)	3 (0.2)	13 (0.7)	19 (1.1)	111 (6.2)	14 (0.8)	42 (2.3)	10 (0.6)	13 (0.7)	5 (0.3)	129 (7.2)	5 (0.3)	35 (1.9)	24 (1.3)	5 (0.3)	7 (0.4)	566 (31.5)
Total	40 (2.3)	26 (1.5)	265 (14.8)	15 (0.8)	87 (4.9)	26 (1.5)	739 (41.2)	27 (1.5)	112 (6.2)	26 (1.5)	22 (1.2)	26 (1.5)	211 (11.8)	5 (0.3)	60 (3.3)	57 (3.2)	42 (2.3)	9 (0.5)	1,795 (100.0)
Sensitivity (%)	75.0	61.5	58.1	80.0	85.1	26.9	85.0	48.2	62.5	61.5	40.9	80.8	38.9	0.0	41.7	57.9	88.1	22.2	

UCOD, underlying cause of death; ICR, Iowa Cancer Registry; IBD, intrahepatic bile ducts; MN, malignant neoplasm; NOS, not otherwise specified; ONS, other nervous system.

TABLE III. Cross-Classification of All Conditions Coded on Death Certificate and Site of First Incident Primary Malignant Neoplasm Reported by the ICR

Cancers identified among all conditions coded on death certificates	First malignant primary tumor site identified by Iowa Cancer Registry (frequency (%))																		
	Esophagus	Stomach	Colon and rectum	Liver and IBD	Pancreas	Larynx	Lung and bronchus	Skin melanoma and other MN	Breast	Cervix uteri	Corpus and uterus, NOS	Ovary and other uterine adnexa	Prostate	Testis	Urinary bladder	Kidney and renal pelvis	Brain and ONS	Thyroid	Total
Esophagus	34 (1.9)																		34 (1.9)
Stomach		17 (1.0)																	17 (1.0)
Colon and rectum			178 (9.9)																178 (9.9)
Liver and IBD				12 (0.7)															12 (0.7)
Pancreas					76 (4.2)														76 (4.2)
Larynx						9 (0.5)													9 (0.5)
Lung and bronchus							665 (37.1)												665 (37.1)
Skin melanoma and other MN								14 (0.8)											14 (0.8)
Breast									79 (4.4)										79 (4.4)
Cervix uteri										16 (0.9)									16 (0.9)
Corpus and uterus, NOS											9 (0.5)								9 (0.5)
Ovary and other uterine adnexa												22 (1.2)							22 (1.2)
Prostate													109 (6.1)						109 (6.1)
Testis														0 (0)					0
Urinary bladder															27 (1.5)				27 (1.501)
Kidney and renal pelvis																37 (2.1)			37 (2.1)
Brain and ONS																	37 (2.1)		37 (2.1)
Thyroid																		2 (0.1)	2 (0.1)
ICR site code not among all conditions coded (Under-reported)	6 (0.3)	9 (0.5)	87 (4.9)	3 (0.2)	11 (0.6)	17 (1.0)	74 (4.1)	13 (0.7)	33 (1.8)	10 (0.6)	13 (0.7)	4 (0.2)	102 (5.7)	5 (0.3)	33 (1.8)	20 (1.1)	5 (0.3)	7 (0.4)	452 (25.2)
Total	40 (2.2)	26 (1.5)	265 (14.8)	15 (0.8)	87 (4.9)	26 (1.5)	739 (41.2)	27 (1.5)	112 (6.2)	26 (1.5)	22 (1.2)	26 (1.5)	211 (11.8)	5 (0.3)	60 (3.3)	57 (3.2)	42 (2.3)	9 (0.50)	1,795 (100.0)
Sensitivity (%)	85.0	65.4	67.2	90.0	87.4	34.6	90.0	51.9	70.5	61.5	40.9	84.6	51.7	0.0	45.0	64.9	88.1	22.2	

UCOD, underlying cause of death; ICR, Iowa Cancer Registry; IBD, intrahepatic bile ducts; MN, malignant neoplasm; NOS, not otherwise specified; ONS, other nervous system.

ICR. Use of all conditions coded, which included up to 14 causes and contributing factors, decreased under-reporting to 452 (25.2%) of the 1,795 ICR-identified neoplasms.

As expected, site-specific under-reporting also improved with the use of all conditions coded, compared to the UCOD only, for 12 of the 18 cancer sites (Table IV). Since all conditions coded included the UCOD, increased under-reporting was not possible. Under-reporting by all conditions coded on the death certificate ranged from 10% (74/739) for cancer of the lung and bronchus to 100% (5/5) for testicular cancer. PV+ values of the UCOD were 80% or greater for all cancer sites, except cancer of the corpus and uterus NOS (69%) and an indeterminable result for testicular cancer. *P*-values less than 0.05 indicated that when using all conditions coded, differential misclassification occurred for 10 cancer sites. Using the UCOD only, under-reporting was significant for three additional neoplasms.

Multivariate logistic regression modeling indicated that survival time and age at diagnosis were the only statistically significant predictors, at the $\alpha = 0.05$ level, of matching COD and ICR site code in both final models (Tables V and VI). The adjusted odds of matching UCOD to ICR cancer site decreased by an estimated 19% for every 1-year increase in survival time (*P*-value <0.001) (Table V). Younger age at diagnosis had significantly greater odds of agreement compared to the 75+ age group (*P*-value <0.001). Male workers had lower odds of agreement (*P*-value = 0.008) compared to females. Odds ratios for workers diagnosed during 1973–1983 and 1984–1994 compared to workers diagnosed during 1995–2005 were lower, but the differences were not statistically significant.

For all conditions coded a 1-year increase in survival time decreased odds of agreement with the ICR cancer incidence data by an estimated 18% (*P*-value <0.001) (Table VI). Similar to UCOD, younger ages at diagnosis had significantly greater odds of agreement compared to the 75+ age group (*P*-value <0.001). Male workers had lower odds of agreement compared to females, but the difference was not statistically significant.

DISCUSSION

This study compared the use of all conditions coded on the death certificate to the more traditional approach, using the UCOD only, to reflect the site of the first primary cancer identified by the ICR among a population of former DOD contract workers. As hypothesized, death certificates under-reported cancer incidence, but for 12 of the 18 cancer sites studied, use of all conditions coded improved ascertainment compared to the UCOD only.

TABLE IV. Percent Under-Reported by All Conditions Coded and UCOD and Predictive Value Positive of UCOD Relative to the Site of the First Incident Primary Cancer Reported by the ICR

Site	n	% Under-reported by all conditions coded	*P*-value[a]	% Under-reported by UCOD	PV+ of UCOD	*P*-value[a]
Lung and bronchus	739	10.0	<0.001	15.0	98.9	<0.001
Brain and ONS	42	11.9	1.000	11.9	90.2	1.000
Pancreas	87	12.6	0.057	14.9	96.1	0.021
Esophagus	40	15.0	0.754	25.0	88.2	0.180
Ovary and other uterine adnexa	26	15.4	0.125	19.2	100	0.063
Liver and IBD	15	20.0	1.000	20.0	80.0	1.000
Breast	112	29.5	<0.001	37.5	100	<0.001
Colon and rectum	265	32.8	<0.001	41.9	96.9	<0.001
Stomach	26	34.6	0.065	38.5	88.9	0.039
Kidney and renal pelvis	57	35.1	0.004	42.1	89.2	<0.001
Cervix uteri	26	38.5	0.002	38.5	100	0.002
Skin melanoma and other MN	27	48.1	0.002	51.8	92.9	<0.001
Prostate	211	48.3	<0.001	61.1	100	<0.001
Urinary bladder	60	55.0	<0.001	58.3	86.2	<0.001
Corpus and uterus NOS	22	59.1	0.167	59.1	69.2	0.049
Larynx	26	65.4	<0.001	73.1	100	<0.001
Thyroid	9	77.8	0.016	77.8	100	0.016
Testis	5	100.0	0.063	100.0	Und[b]	0.063

UCOD, underlying cause of death; ICR, Iowa Cancer Registry; ONS, other nervous system; IBD, intrahepatic bile ducts; MN, malignant neoplasm; NOS, not otherwise specified.
[a]*P*-values calculated using two-sided exact binomial test.
[b]Undefinedno cases identified as UCOD.

TABLE V. Odds Ratio of Agreement Between UCOD and ICR Incidence Data

Variable	OR[a] (95% CI)	*P*-value
Survival time (1-year increase)	0.81 (0.79-0.84)	<0.001
Age at diagnosis (years)		<0.001
22–60	3.12 (2.26–4.32)	
61–68	2.58 (1.89–3.53)	
69–74	1.25 (0.93–1.66)	
75+	1.00 (referent)	
Sex		0.008
Male	0.72 (0.56–0.92)	
Female	1.00 (referent)	
Diagnosis era		0.372
1973–1983	0.82 (0.60–1.11)	
1984–1994	0.86 (0.66–1.12)	
1995–2005	1.00 (referent)	

UCOD, underlying cause of death; ICR, Iowa Cancer Registry.
[a]Odds ratio adjusted for all other variables in the table using logistic regression.

For the six cancer sites that showed the best agreement between death certificate and ICR data (cancers of the lung and bronchus, brain and ONS, pancreas, esophagus, ovary and other uterine adnexa, and liver and IBD), all conditions coded under-reported 10–20% of incidence, whereas the UCOD failed to detect 11.9–25% of cases identified by the ICR. In 2012, four cancer sites, neoplasms of the prostate, breast, lung and bronchus, and colon and rectum, were predicted to represent over 51% of all incident cancers that were registry-identified in the United States [Howlader et al., 2012]. The same four sites accounted for 74% of the incident cancers identified by the ICR for this cohort. Under-reporting by all conditions coded for these four common neoplasms ranged from 10% for cancer of the lung and bronchus to 48.3% for prostate cancer. The UCOD failed to detect 15–61.1% of the same cancers. Differential misclassification was statistically significant for all four cancer sites. PV+ values indicated that most cancers listed as the UCOD were also identified by the ICR.

TABLE VI. Odds Ratios of Agreement Between All Conditions Coded and ICR Incidence Data

Variable	OR[a] (95% CI)	*P*-value
Survival time (1-year increase)	0.82 (0.80–0.84)	<0.001
Age at diagnosis (years)		<0.001
22–60	2.40 (1.73–3.34)	
61–68	2.16 (1.55–3.01)	
69–74	1.12 (0.82–1.52)	
75+	1.00 (referent)	
Sex		0.094
Male	0.81 (0.62–1.05)	
Female	1.00 (referent)	

ICR, Iowa Cancer Registry.
[a]Odds ratio adjusted for all other variables in the table using logistic regression.

The exact binomial test indicated that when using all conditions coded, under-reporting was statistically significant for ten cancer sites. When using the UCOD only, differential misclassification occurred for three additional cancers.

The misclassification was likely due to a complex interplay of numerous factors. These include the survival time after development of cancer, the tendency of certain primary cancer types to develop metastases in specific organs [Chambers et al., 2002; Fokas et al., 2007] that have increased likelihood of being reported on the death certificate (e.g., colorectal cancers often metastasize to the lung or liver), cancer-related illnesses (e.g., lung cancer and pneumonia), and the severity of comorbid conditions (e.g., heart disease, pulmonary disease, etc.).

Additionally, the small sample sizes for some cancer types in this study made it difficult to accurately estimate measures of agreement for death certificate data relative to ICR-reported cancer incidence. For example, with only five incident cases of testicular cancer recorded by the ICR, there was insufficient power to reject the null hypothesis, despite the fact that death certificates under-reported every case of testicular cancer among the IAAAP worker cohort. But for frequently occurring incident cancers, this measure likely provided a reliable estimate of how well the death certificate reflects cancer incidence. The use of all conditions coded, as compared to the more typical UCOD only, may improve estimates of under-reporting for similar cohorts not covered by central cancer registries.

Multivariate logistic regression modeling was performed to examine the effect of survival time, age at diagnosis, sex, diagnosis era, ICD revision in use at time of death, and race on the odds of agreement between UCOD or all conditions coded and ICR site code. As hypothesized, the odds of agreement lessened significantly with increasing survival time after cancer diagnosis for both UCOD and all conditions coded compared to the ICR incidence data. Both models also indicated that age at diagnosis was a significant predictor of agreement between death certificates and the ICR data. Odds ratios were increased for workers whose age at diagnosis was less than or equal to the median for the cohort compared to workers in the oldest group. Males and workers diagnosed in earlier time frames were found to have decreased odds of agreement as compared to females and those in the most recent diagnosis era, respectively.

Demers et al. [1992] calculated similar estimates of relative risk for rapidly fatal cancers from both tumor registry and mortality data, but for cancers with better survival, more cases were identified using incidence data as compared to death certificates. The authors noted that additional cancers were identified on death certificates, but only cancers listed as the UCOD were included in their analyses. Similarly, our findings indicated that death certificates underestimated cancer incidence. Rather than focus on specific cancers with varying mortality rates, we evaluated survival time for all workers. We found survival time to be a main effect in the logistic regression models to estimate the odds of agreement between death certificates and the ICR incidence data.

In two separate analyses of data, including all conditions coded on the death certificate, from the U.S. Radiological Technologists Study researchers reported an overall 35% under-ascertainment of cancers identified by death certificates compared to registry data and that 29.4% of all cancers included in the study were identified by death records [Sigurdson et al., 2003; Freedman et al., 2006]. Limitations of these studies may include lack of both site-specific under-reporting figures and verification of cancers reported on death certificates only. The authors also point out that this cohort may not be comparable to non-healthcare workers due to their access to health care, medical knowledge, and heightened awareness of their potential exposure and risk of adverse health outcomes. Our study also found that death certificates under-reported cancer incidence compared to the ICR data.

Two occupational studies, Cottrell et al. [1992] and Davis et al. [1992], reported better disease surveillance results using manual review of death certificates to ascertain any mention of the disease of interest. Our study employed a similar methodology by using all conditions coded on the death certificate and found more complete ascertainment of cancer occurrence for 12 of the 18 diseases studied.

The findings of our research support the views of previous authors [Feinstein and Esdaile, 1987; Boyle, 1989], who discussed the limitations of using death certificates for ascertainment of cancer incidence data. The findings also agree with other researchers [Demers et al., 1992; Pickle et al., 2007] who found that longer survival time was a negative predictor of agreement between death certificate data and SEER registry incidence data. However, use of all conditions coded on the death certificate should improve ascertainment of cancer occurrences compared to the UCOD alone.

Results of this study should be generalizable to other occupational cohorts in similar workplaces and time frames. Measures of agreement between death certificate data and cancer incidence data indicate that, particularly in areas not covered by population-based tumor registries, death certificates can serve as reasonable detectors of cancer-specific incidence for certain neoplasms, especially rapidly fatal cancers. In addition, age at diagnosis was found to be a significant predictor of agreement between death certificate and cancer incidence data. Differential misclassification by both all conditions coded and the UCOD alone was noted for most cancers. Findings for diseases with small sample sizes should be interpreted carefully as measures of agreement for such diseases are difficult to estimate accurately.

Primary among the strengths of this study was the overall high quality of the data from the ICR. This study also evaluated a novel approach for identifying cancer occurrence from death certificates by comparing the ability of all conditions coded versus the UCOD to reflect cancer incidence.

Study limitations included cancer sites being identified at the three-digit ICD site code level, which eliminated from our evaluation non-Hodgkin lymphoma, one of the 10 most common cancers, as well as other important neoplasms. The sample size also limited our ability to adequately evaluate less frequently occurring cancers. It is uncertain whether these results are generalizable to other populations. About one in six U.S. cancer patients are afflicted with multiple primary neoplasms [Grouse, 2006]. Workers with multiple primaries, approximately 15.3% of the potential cohort, were excluded from this study. Future work includes developing methods to include subjects with multiple primary malignancies. Additional future investigation may include development of methods to predict which of multiple cancers listed on the death certificate is most likely to correctly identify the first primary.

SUMMARY

For over 50 years, the American Cancer Society (ACS) has used death certificate data as part of its formula for estimating cancer incidence, and periodic reviews have deemed the results useful and reliable [Silverberg and Lubera, 1983]. When tumor registry data are not available, our findings indicate that death certificates can be useful indicators of cancer incidence. This research also quantified the extent to which death certificates under-reported cancer incidence for the 18 eligible cancer sites. Additionally, use of all conditions coded on the death certificate was shown to decrease the extent of under-reporting, compared to the more commonly utilized UCOD, for many of the cancer sites.

Survival time after cancer diagnosis and age at diagnosis were found to be significant predictors of the odds that death certificate and cancer incidence data would agree. Workers with shorter survival, younger age at diagnosis, and female sex all were found to have increased odds of agreement compared to their counterparts.

Assuming that the findings of this article are generalizable to other worker cohorts, the results may be useful for estimating the degree of under-reporting in other studies that rely on death certificates to reconstruct cancer incidence. However, these findings are most generalizable to worker cohorts in the upper Midwest. Similar studies are called for in other areas of the United States to further assess the representativeness of these findings.

REFERENCES

Bedford RL. 2009. Utility of death certificate data in predicting cancer incidence. IA: The University of Iowa.

Boyle P. 1989. Relative value of incidence and mortality data in cancer research. Recent Results Cancer Res 114:41–63.

Chambers AF, Groom AC, MacDonald IC. 2002. Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer 2(8):563–572.

Cottrell A, Schwartz E, Sokas R, Kofie V, Welch L. 1992. Surveillance of sentinel occupational mortality in the District of Columbia: 1980 to 1987. Am J Public Health 82(1):117–119.

Davis LK, Martin TR, Kligler B. 1992. Use of death certificates for mesothelioma surveillance. Public Health Rep 107(4):481–483.

Demers PA, Vaughan TL, Checkoway H, Weiss NS, Heyer NJ, Rosenstock L. 1992. Cancer identification using a tumor registry versus death certificates in occupational cohort studies in the United States. Am J Epidemiol 136(10):1232–1240.

Feinstein AR, Esdaile JM. 1987. Incidence, prevalence, and evidence. Scientific problems in epidemiologic statistics for the occurrence of cancer. Am J Med 82(1):113–123.

Field RW, Smith BJ, Platz CE, Robinson RA, Neuberger JS, Brus CP, Lynch CF. 2004. Lung cancer histologic type in the surveillance, epidemiology, and end results registry versus independent review. J Natl Cancer Inst 96(14):1105–1107.

Fokas E, Engenhart-Cabillic R, Daniilidis K, Rose F, An HX. 2007. Metastasis: The seed and soil theory gains identity. Cancer Metastasis Rev 26(3–4):705–715.

Freedman DM, Sigurdson AJ, Doody MM, Love-Schnur S, Linet MS. 2006. Comparison between cancers identified by state cancer registry, self-report, and death certificate in a prospective cohort study of US radiologic technologists. Int J Epidemiol 35(2):495–497.

Grouse L. 2006. Second cancers complicate long-term survival [Internet]: National Cancer Institute; c2006 [cited 2009/12/02]. Available from: http://benchmarks.cancer.gov/2006/04/second-cancers-complicate-long-term-survival/

Hetzel AM. 2008. History and organization of the vital statistics system [Internet]. Hyattsville, MD: National Center for Health Statistics. c1997 [cited 2008/03/11]. Available from: http://www.cdc.gov/nchs/data/misc/usvss.pdf

Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ. Cronin KA, editors. 2012. SEER Cancer Statistics Review 1975–2009 (Vintage 2009 Populations) [Internet]; c2012 [cited 2013/01/03]. Available from: http://seer.cancer.gov/csr/1975_2009_pops09/

Moriyama IM, Loy RM, Robb-Smith AHT. 2011. History of the statistical classification of diseases and causes of death [Internet]. Hyattsville, MD: National Center for Health Statistics (U.S.). c2011 [cited 2013/02/09]. Available from: http://www.cdc.gov/nchs/data/misc/classification_diseases2011.pdf

National Cancer Institute. 2008. Overview of the SEER Program [Internet]: National Cancer Institute [cited 2008/02/27]. Available from: http://seer.cancer.gov/about/

National Center for Health Statistics. 2003. U.S. Standard Certificate of Death [Internet]. Hyattsville, MD: U.S. Centers for Disease Control and Prevention, National Center for Health Statistics; c2003 [cited 2008/02/02]. Available from: http://www.cdc.gov/nchs/data/dvs/DEATH11-03final-ACC.pdf

NPCR. 2008. About the Program [Internet]. Atlanta, GA: U. S. Centers for Disease Control and Prevention; c2008 [cited 2008/02/26]. Available from: http://www.cdc.gov/cancer/npcr/about.htm

National Program of Cancer Registries. [NPCR]. 2009. Registry Plus™ Link Plus [Internet]; c2009 [cited 2013/05/30]. Available from: http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm

Percy C, Stanek E III, Gloeckler L. 1981. Accuracy of cancer death certificates and its effect on cancer mortality statistics. Am J Public Health 71(3):242–250.

Percy CL, Miller BA, Gloeckler Ries LA. 1990. Effect of changes in cancer classification and the accuracy of cancer death certificates on trends in cancer mortality. Ann N Y Acad Sci 609 87–97; discussion 97–99.

Pickle LW, Hao Y, Jemal A, Zou Z, Tiwari RC, Ward E, Hachey M, Howe HL, Feuer EJ. 2007. A new method of estimating United States and state-level cancer incidence counts for the current calendar year. CA Cancer J Clin 57(1):30–42.

Sigurdson AJ, Doody MM, Rao RS, Freedman DM, Alexander BH, Hauptmann M, Mohan AK, Yoshinaga S, Hill DA, Tarone R, Mabuchi K, Ron E, Linet M. 2003. Cancer incidence in the US radiologic technologists health study, 1983–1998. Cancer 97(12):3080–3089.

Silverberg E, Lubera JA. 1983. A review of American cancer society estimates of cancer cases and deaths. CA Cancer J Clin 33(1):2–8.

West MM, Lynch CF, McKeen KM, Olson DB, Platz CE, Field RW, Weiner G. 2012. 2012 Cancer In Iowa [Internet]; c2012 [cited 2013/01/03]. Available from: http://www.public-health.uiowa.edu/shri/pubs/pdf/Cancer_2012.pdf

WHO. 2000. International Classification of Diseases for Oncology, 3rd edition (ICD-O-3) [Internet]. Lyon, France: WHO. c2000 [cited 2008/05/30]. Available from: http://www.who.int/classifications/icd/adaptations/oncology/en/

World Health Organization, [WHO]. 2008. History of the development of the ICD [Internet]. Geneva, Switzerland: World Health Organization. c2008 [cited 2008/02/03]. Available from: http://www.who.int/classifications/icd/en/HistoryOfICD.pdf

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.