Dartmouth GEISEL SCHOOL OF MEDICINE

Using NAACCR Survival Tables to Look at Deaths

Bruce Riddle PhD, Scot Zens PhD, Maria Celaya MPH, CTR, Judy Rees BM, BCh, MPH, PhD
Geisel School of Medicine, Dartmouth College

Introduction

Annual death clearance, even with linkage to the National Death Index, is hindered by practical issues including missing social security number and other identifiers that may lead to errors in ascertaining deaths. Missing death information could be investigated for selected patients whose survival seems unusually good. The NAACCR Survival Analysis Task Force tables offer an opportunity to take a new look at deaths because the 5-year age-standardized relative survival ratios (RSR) indicate the relative risk of death under specific circumstances such as cancer site, stage and patient age.

When the RSR data suggest that, on average, there is a low probability of survival in a patient who appears to be alive in the registry database, possible explanations include unusually good survival, or flaws in death ascertainment. If we can use statistical methods to identify survival anomalies (patients we expect to have died but who have no death information in the registry), we can check vital status through methods other than death certificate linkage. In this study, we investigated survival anomalies in New Hampshire's cancer registry data.

Table 1. Explanatory variables tested in the model.

	Age <65 N=21,927	Age 65+ N=22,754
Died	N (%) 6142 (28%)	N (%) 12,948 (56.9%)
Age at diagnosis, mean (SD)	52.4 (10.51)	75.3 (7.21)
Years since diagnosis, mean (SD)	7.8 (2.02)	7.8 (2.02)
Stage at diagnosis*:		
Local	53.1% (11,649)	48.3% (10991)
Regional	23.7% (5192)	19.6% (4456)
Distant	20.9% (4577)	26.3% (5979)
Race:		
White	96.4% (21151)	97.7% (22223)
Non white	3.5% (778)	2.3% (531)
Marital status**		
Married	64.5% (14152)	55.9% (12723)
Single	15.0% (33013% (677)	6.2% (1421)
Widowed		24.0% (5467)
Primary Payer at Diagnosis		
Not insured	5.4% (1191)	0.3% (64)
Insurance NOS	15.9% (3480)	2.7% (613)
Private	59.6% (13090)	6.6% (1497)
Medicare	7.8% (1722)	81.7% (18597)
Medicaid	4.7% (1022)	0.5% (124)
RSR by site and sex***	71.1 (28.0)	61.0 (34.8)
RSR by stage and sex***	72.3 (28.0)	
RSR by age and sex***	69.7 (32.7)	

*Stage based on Derived Summary State 2000
**(Aizer, 2013) used SEER data 2004-2008 and stated as with metastatic cancer, under treatment and death resulting from their cancer."
***5-Year Age-Standardized Relative Survival Ratios (RSR) for Cancers Diagnosed 2005-2011, Complete Method, Follow-Up Through 2011, NAACCR U.S. Registries, By Primary Site Category and Registry, Whites

Results

In those aged under 65, 6142 (28%) of 21,927 patients had died. The average age at diagnosis was 52.4 with a median age of 55. In the multivariable model (Table 2), lower odds of death were associated with private insurance, being married, and having a cancer with lower age-specific RSR and stage-specific RSR. Increasing age, the number of years since diagnosis, and distant disease at diagnosis were associated with greater odds of death. Insurance status also significantly affected survival.

In those aged over 65, 12,948 (56.9%) of 22,754 patients had died. The average age at diagnosis was 75.2 with a median of 74. Regional disease, lower stage-specific RSR and being married are all protective, while odds of death increased with increasing age and years since diagnosis.

Table 2: Multivariate models of factors associated with death (odds ratio [OR], 95% confidence interval)

	Model for <65 years		Model for 65+ years	
Variables	OR (95% CI)	P	OR (95% CI)	P
Age at diagnosis	1.42 (1.04-1.05	<.0001	1.11 (1.10-1.11)	<.0001
Years since diagnosis	1.20 (1.18-1.23)	<.0001	1.33 (1.30-1.35)	<.0001
Marital status married	0.82 (0.75-0.90)	<.0001	0.73 (0.68-0.78)	<.0001
Race white	1.52 (1.21-1.92)	.0003		
Stage: regional			0.86 (0.79-0.93)	<.0001
Stage: distant	1.19 (1.05-1.34)	.0050		
RSR by stage and sex	0.95 (0.95-0.95)	<.0001	0.95 (0.95-0.95)	<.0001
RSR by site and sex	1.01 (1.00-1.01)	<.0001		
RSR by age and sex	0.99 (0.99-0.99)	<.0001		
Primary Payer: not insured	1.40 (1.13-1.73)	.0024		
Primary Payer: insurance NOS	0.66 (0.55-0.79)	<.0001		
Primary Payer: private	0.62 (0.52-0.72)	<.0001		
Primary Payer: medicaid	1.82 (1.45-2.28)	<.0001		
Primary Payer: medicare	1.34 (1.10-1.63)	.0038		

Bring out your dead!
Bring out your dead!

Table 3: Under 65.

	Registry shows patient alive Total alive: N = 15,785		Registry shows patient deceased Total dead: N = 6142	
	"Should be dead"	All "alive"	"Should be alive"	All "deceased"
Selected cases, N	257 (1.6%)	15,785 (100%)	1065 (17.3%)	6142 (100%)
Mean age (SD)	56.6 (7.5)	51.4 (11.0)	52.03 (11.1)	54.7 (8.6)
Percent married	72%	67.3%	63%	57.5%
Private insurance	36%	64.2%	58%	48%
Years since diagnosis, mean (SD)	8.0 (2.03)	7.7 (2.0)	8.0 (2.0)	8.0 (2.0)
Most frequent sites	Lung 41.2% Kidney 5.7% Liver 4.4% Esophagus 3.6%	Lung 26.5% Bone marrow 5.5% Prostate 3.6% Kidney 3.3%	Breast 32.9% Prostate 16.1% Kidney 6.7% Endometrium 4.2%	Lung 26.5% Bone marrow 4.7% Prostate 3.6% Kidney 3.3%

Under 65 Model: Should be Dead
Of a total of 15,785 cancer survivors in the registry, we examined 257 (1.6%), who were least likely to be alive according to the multivariable model. The median age was 56.6. There were on average 35 cases per year, 2005-2011, equally split between male and female. In this subset 56.50% (139) were married, 71.66% (177) had distant disease and 36% (90) had private insurance. Looking at primary site, 41.29% (102) had a lung primary with 22.27% (55) lung upper lobe. After lung, the next three most frequent sites were kidney (5.67%, 14 cases), liver (4.45%, 11 cases), and esophagus (3.65%, 9 cases). The distribution of cases by reporting facility seem to match the routine reporting patterns.

Under 65 Model: Should be Alive
We examined 1065 (17.3%) cases out of 6142 dead, who had the poorest survival. The median age was 55 and a mean of 52. There were, on average, 152 cases per year, 2005-2011, with 54.0% female. In this subset, 63.1% (671) were married and 58.4% (622) had private insurance; another 16.8% (179) had some insurance, NOS. Looking at derived stage, 59.4% (633) were localized and another 16.4% (175) regional, regional lymph nodes only. Looking at primary site, the largest specified group was prostate at 16.1% (172). However, collapsing all the breast sites give 32.9% of cases (350). The next two largest groups were kidney at 6.7% (71) and endometrium 4.2% (45 cases).

These cases have a cause of death. The cause of death matched the primary site at the 3 character level (i.e., 'C02' = 'C02') 39% (425) of the time. For the subset, 74.0% (788) of cases had a cancer as cause of death. The next largest groups were circulation (I), 8.1% (86), respiratory (J) 3.1% (33), digestive (K) 2.3% (25), and endocrine (E) 2.16% (23).

Table 4: Age 65 and older.

	Registry shows patient alive Total alive: N = 9806		Registry shows patient deceased Total dead: N = 12,498	
	"Should be dead"	All "alive"	"Should be alive"	All "deceased"
Selected cases, N	568 (5.8%)	9806 (100%)	516 (4.1%)	12,948 (100%)
Mean age (SD)	78 (7.27)	72.8 (6.0)	69.8 (3.9)	77.0 (7.4)
Years since diagnosis, mean (SD)	8.0 (2.0)	7.39 (2.0)	6.9 (1.6)	8.1 (2.0)
Married	46.4%	63.8%	74.2%	50.0%
Widowed	34.4%	17.0%	9.5%	29.4%
Medicare	82.6%	80.3%	75.4%	82.7%
Distant stage	44.5%	26.3%	0.5%	39.7%
Localized	14.8%	48.3%	83.3%	30.4%
Frequent sites	Lung 33.8% Bone marrow 11.4% Prostate 4.0%	Prostate 27.7% Breast 12.1% Endometrium 3.8% Bone marrow 3.7%	Prostate 49.4% Breast 19.9% Endometrium 3.9%	Lung 26.8% Prostate 7.5% Bone marrow 5.4% Pancreas 2.8%

Over 65 Model: Should be dead
The group contained 568 cases out of 9806 alive. The mean and median age are both 78. There are more females at 53.3% (303) versus males 46.5% (264). In this subset, 46.4% (263) were married and 34.4% (195) widowed. As is expected, 82.57% were on Medicare. For derived summary stage, 44.54% (253) were distant and 32.5% (185) were regional. For primary site, 18.3% (104) were upper lobe lung; adding all the lung site totaled 33.8% (192). The next largest groups were bone marrow 11.4% (65), prostate 4.0% (23) and esophagus 3.3% (19).

Over 65 Model: Should Be Alive or Should not be Dead
This model contained 516 cases out of 12,498 dead. The distribution across years varied from 9 cases in 2005 to 113 in 2011. For this subset, 67.2% (347) were male, 74.5% (383) were married, 75.0% were on Medicare and 11.0% on private insurance. Looking at derived stage, 83.3% had localized disease. The mean for years since diagnosis was 6.9 and median 6.8. For primary site, 49.4% of cases (255) were prostate, and 19.9% (103) were breast. The next largest group were endometrium 3.9% (20).

This group has cause of death. The cause of death matched the primary site at the 3 character level (i.e., 'C02' = 'C02') 18% (93) of the time. For the subset 52.3% (270) had a cancer as cause of death. The next largest group was circulatory system (I) 17.0% (88), respiratory system (J) 7.7% (40), endocrine 5.8% (30). Unlike above, the causes of death were more granular. It takes 15 different classifications to reach 50% of cases. The largest four groups covering just 25% of the cases are lung 8.1% (42), prostate (6.8%) (35), breast 6.4% (33) and atherosclerotic heart disease 3.7% (19).

Methods

We used the RSRs by site, 2005-2011, for the New Hampshire State Cancer Registry and incorporated them into a logistic regression model. We identified malignant cancer cases in adults from the New Hampshire State Cancer Registry, diagnosed 2005-2011, and excluded cases identified only through pathology reports or death clearance, or those with unknown class of case. We restricted the analysis to the 24 sites for which RSR was reported by the NAACCR Survival Analysis Task Force. The analytic dataset was created on February 3, 2016.

Models were created to describe vital status, one each for younger and older patients with a cut-off of 65 years to account for differences in Insurance. In each case we tested a full model with the pre-specified explanatory variables listed in Table 1, and sequentially removed variables with p>0.05 until a final model was obtained. Proc Logistic in SAS 9.4 was used to do the analysis.

Each patient was assigned a predictive value for death according to the model. Extreme values were examined to identify the characteristics of patients whose survival was surprisingly short or long after diagnosis. Arbitrarily, we looked at predictive values of 0.80 and above for survivors expected to be dead and 0.20 or below for those who had died but were expected to be alive.

Discussion

It is unclear how often errors in vital status ascertainment affect registry data. Two subsets of patients were identified, whose vital status was least consistent with the multivariable model. The first subset may contain a mix of individuals who truly had unusually good survival, or cases where the registry have failed to ascertain the death. The second subset either had unusually poor survival, or the death may have been wrongly attributed to the patient on the date specified. Use of the model to identify outliers offers a potential strategy to investigate and improve data quality. In addition, this exercise gives us an opportunity to describe the causes of death in those who do unusually poorly after a cancer diagnosis.

To fully understand the interplay of death and cancer, all the underlying causes of death need to be added to the NAACCR record. In some instances, this will not be possible. However, to understand the "Should Be Alive/Should Not Be Dead" group, we need to better understand when a patient has not died from cancer but from something unrelated or only distantly related. The underlying cause of death can often be thought of as the condition that led a person to seek treatment. The CDC completion of death certificates states "An important feature is the reported underlying cause of death determined by the certifying physician and defined as (a) the disease or injury that initiated the train of morbid events leading directly to death, or (b) the circumstances of the accident or violence that produced the fatal injury." The complications of medical care, which are standard parts of a death certificate, are not included in a NAACCR record. While there are well recognized problems with completion of death certificates by providers, a death certificate is the best data we have to look at mortality.

Next steps

We will try to track down those who seem to have unexpectedly good survival to see whether there has been an error in ascertainment of vital status. The internet has various resources to identify whether there has been a funeral, or whether the patient is in fact still alive. Individuals whose missing or changed data preclude a successful linkage with death files may be tracked down through these alternate means.

Cited Works:

Aizer AA, Chen M, McCarty EP, Mendu ML, Koo,S, Wilhite TJ, Graham PL, Choueiri TK, Hoffman KE, Martin NE, HU JC, Nguyen PL. Marital Status and Survival in Patients With Cancer. Journal of Clinical Oncology (2013) 31:3869-3876.

Johnson CJ, Weir HK, Mariotto AB, Nishri D, Wilson R (eds). Cancer in North America: 2008-2012 Volume Four: Cancer Survival in the United States and Canada 2005-2011. Springfield, IL: North American Association of Central Cancer Registries, Inc. March 2016.

Physicians' Handbook on Medical Certification of Death, 2003 Revision, Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statisticds, Hyattsville, Maryland, April 2003, DHHS Publication No. (PHS) 2003-1108

Acknowlegements

This project was supported in part by the Centers for Disease Control and Prevention's National Program of Cancer Registries, cooperative agreement 5 NU58DP003930-04-00 awarded to the New Hampshire Department of Health and Human Services, Division of Public Health Services, Bureau of Public Health Statistics and Informatics, Office of Health Statistics and Data Management. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or New Hampshire Department of Health and Human Services.