NIH Public Access
Author Manuscript
Cancer. Author manuscript; available in PMC 2014 May 15.

Published in final edited form as:
Cancer. 2013 May 15; 119(10): 1900–1907. doi:10.1002/cncr.27968.

Assessing the Utility of Cancer-Registry-Processed Cause of Death in Calculating Cancer-Specific Survival

Chung-Yuan Hu, M.P.H., Ph.D., **Yan Xing, MD, MS, Ph.D**, **Janice. N. Cormier, M.D, M.P.H**, and **George J. Chang, M.D., M.S.**
Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center

Abstract

Background—Cancer registries use algorithms to process causes of death (COD) from death certificate but uncertainties remain in its accuracy and utility in calculating cancer-specific survival (CSS).While it is impractical to reconfirm the COD through primary medical record review, one could alternatively compare the observed cancer deaths with the number of attributed deaths as estimated by relative survival (RS) approach to determine its utility in CSS estimation.

Method—Six major cancer types were evaluated using the Surveillance Epidemiology and End Results data (1988-1999 cohort). The COD utility was quantified by the observed-to-expected ratio (O/E ratio) approach, calculated as the SEER-documented observed number of cancer-specific deaths divided by the number of expected deaths attributed to the malignancies through RS approach. Favorable utility would have an O/E ratio close to 1.

Results—We identified 338,445 subjects and their O/E ratios were 0.97, 0.98, 0.90, 1.07, 1.02, and 0.92 for breast, colorectal, lung, melanoma, prostate, and pancreas cancer, respectively. O/E ratios varied slightly with patients' age, race and tumor stage, but not by sex. CSS for lung cancer appeared to be overestimated considerably. Patients with multiple cancer diagnoses had poor O/E ratio compared to patients with only one cancer.

Conclusions—The utility of COD in calculating CSS is varying dependent upon the risk of cancer-related mortality and non-tumor factors. However, the impact of this variation on CSS was generally small. The COD as assigned by cancer registries has acceptable validity and CSS is considered an acceptable surrogate for RS in most circumstances.

Keywords

cause of death; SEER; cancer-specific survival; relative survival; cancer

Background

The Surveillance, Epidemiology, and End Results (SEER) Program[1] of the National Cancer Institute (NCI) provides information on cancer incidence and survival statistics that are largely applied in daily medical practice. SEER also reports cause of death (COD) as determined by cancer registries using pre-defined algorithms to process COD from death certificates in order to identify a single, disease-specific, underlying COD. However,

Corresponding Author: George J. Chang, M.D. M.S. Associate Professor of Surgery TEL: 713-563-1875 gchang@mdanderson.org 1400 Pressler Street, #FCT 17.6060 Houston, TX 77030 .

Competing interests The authors declare that they have no competing interests.

Authors' contributions CYH and GJC conceived and designed the study. CYH and YX inputted and undertook the statistical analysis of the data. CYH drafted the manuscript. GJC and JNC reviewed and revised the manuscript. All authors have read and approved the final version of the paper.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

questions have long been raised concerning the reliability of the COD assignment and its accuracy on cancer-specific survival (CSS) estimation[2-6].

When COD reliability is uncertain, relative survival (RS)[7] is commonly used as an alternative. Although both CSS and RS are classical net survival measurements used to quantify the excess mortality attributable to the disease, CSS differs from RS with a number of advantages and disadvantages. CSS is defined as the proportion of patients alive with a specific disease whereas deaths from causes other than the disease of interest are censored or uncounted in this measurement. RS is defined as the ratio of observed survival to the expected survival in a comparable cohort of general population[8]. The primary advantage using RS is that no COD information is required, thereby bypassing the COD inaccuracy issue and difficulties in outcome definition. However, the RS approach requires detailed life tables for comparable populations that are not always directly available for research applications. Moreover, methods for RS analysis are less well recognized by clinical researchers and not readily employed using common statistical packages. Therefore, CSS may be a more practical and preferred measure of cancer survival statistics in a dataset of which COD information is available.

One way to ensure the reliability of COD is to reconfirm the COD through meticulous review of the primary medical records[9]–impractical for a large dataset like SEER. Prior study has indicated that the RS approach can be sued to obtain the expected number of malignancy-attributable deaths[10]. One can then compare this number of malignancy-attributable deaths to the SEER-documented observed number of cancer-specific deaths to acquire the concordance of these two estimates[11]. In the best case scenario, perfect concordance indicates equivalence of not only the number of cancer-related deaths but also net survival estimates.

The objective of the current study was to evaluate the utility of COD in CSS estimation and the concordance between RS and CSS. We hypothesized that RS and CSS have concordances varied by cancer sites, cancer stage and patient characteristics. Because cancer mortality statistics and survival outcomes derived from SEER are widely applied in medical practice (e.g. to inform treatment decisions and prognosis), it is important to ensure the CSS has acceptable concordance with RS so that one can readily obtain net survival by CSS instead of RS that requires detail life table for matched general population.

Methods

Data source and case identification

The SEER data (April 2010 release) we used contains 13 population-based cancer registries that together collect data for all malignancies diagnosed[1]. SEER routinely collects information on patient demographics, primary tumor site, tumor morphology, disease stage at diagnosis, first course of treatment (radiotherapy and surgery), vital status and COD using the combined methods of passive and active follow-up.

Patients eligible for this study included those with microscopically confirmed adenocarcinoma of the breast (ICD-O-3: C50.0-C50.9 with histology codes 8050, 8140-8147, 8160-8162, 8180-8221, 8250-8507, 8514, 8520-8551, 8560, 8570-8574, 8576, 8940-8941); adenocarcinoma of the colorectum (ICD-O-3: C18.0-C18.9 with histology codes 8140,8210-8211,8220-8221,8260-8263,8470,8480-8481,8490); adenocarcinoma, bronchioloalveolar carcinoma, large cell carcinoma, squamous cell carcinoma, and other non-small cell carcinoma of the lung (ICD-O-3: C34.0-C34.9 with histology codes 8140, 8251, 8255, 8260, 8310, 8323, 8480, 8481, 8570, 8250, 8252, 8253, 8012, 8031, 8052, 8070-8074, 8010, 8020, 8022, 8032, 8033, 8046, 8050, 8490, 8550, and 8560); superficial,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

nodular, acral and other malignant melanoma of the skin (ICD-O-3: C44.0-C63.2 with histology codes 8720-8721, 8723, 8730, 8740-8745, 8770-8772); adenocarcinoma of the prostate (ICD-O-3: C61.9 with histology codes 8010, 8140-8570); or adenocarcinoma of the pancreas (ICD-O-3: C25.0-C25.9 with histology codes 8050, 8140-8147, 8160-8162, 8180-8221, 8250-8507, 8514, 8520-8551, 8560, 8570-8574, 8576, 8940-8941). These cancer sites and histologies were chosen as they represented common malignancies with both low and high underlying cancer-specific mortality. Our study cohort included patients diagnosed from January 1988 through December 1999 to secure all subjects that had at least 8 years follow-up through December 2007. The selected 8-year follow-up was based upon our preliminary finding as illustrated in Figure 1. We noted the relative survival curve was deemed flat beyond 8 years following diagnosis. The plateau of the curve indicates the excess mortality of malignancy is minimized. We can then estimate the total number of deaths attributed the malignancy for the entire 8-year period and compare this number to the number of cancer-specific deaths as documented in the SEER dataset. The O/E ratio can still be calculated using different defined time to plateau. For example, the O/E ratio for colorectal cancer (1988-1999 cases) was 0.979, 0.977 and 0.971 using the 7^{th}, 6^{th}, and 5^{th} years as the defined plateau for O/E ratio calculation, respectively.

We employed SEER tumor size, extent of disease and number of regional nodes to re-stage patients according to the AJCC (American Joint Committee on Cancer) 6^{th} edition[12], except for the breast cancer in which the available elements only permitted 5^{th} edition staging. Common exclusion criteria for SEER-based research were applied, including cases with unknown age, less than 18 or greater than 90 years, or if the cancer reporting source was autopsy, nursing home, hospice, or death certificate.

Statistical Analyses

The SEER*Stat program (version 6.22, National Cancer Institute) was used to obtain both CSS and RS through December 2007. CSS was calculated using the SEER COD Recode variable to define the set of individuals who had died of the cancer (SEER COD Recode 26000 for breast, 21040-21050 for colorectum, 22030 for lung, 25010 for melanoma of the skin, 28010 for prostate, and 21100 for pancreas); cases were defined censored if the death occurred from other causes or if the patient was alive at the time of last follow-up. RS was calculated as the ratio of the observed (overall) survival in the study cohort to the expected survival of the general U.S. population matched on the basis of age, gender, race and single calendar year. Relative survival was calculated using the Ederer I method.

Using the 1988-1999 cohort, the 8-year cumulative number of expected deaths attributed to specific malignancies was calculated by subtracting the 8-year cumulative number of expected deaths in the matched general population as estimated by RS approach from the 8-year cumulative observed (overall) deaths as documented within SEER. The 8-year observed number of cancer-specific deaths as documented within the SEER COD Recode variable was then divided by this number to yield the observed-to-expected ratio (O/E ratio). For example, a ratio less than 1.0 indicates SEER documented a lower-than-expected number of cancer-specific deaths which would result in an overestimated CSS; a ratio greater than 1.0 indicates SEER documented a higher-than-expected number of cancer-specific deaths which would result in an underestimated CSS. The chance variation of the O/E ratio was determined based on the Z-score using the normal approximation to the binomial: $z=\frac{\widehat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$; where $\widehat{p}$ denotes the observed cancer mortality rate (the proportion of patients dying from the specific cancer during the 8-year follow-up); p_0 denotes the expected cancer mortality rate; n denotes the number of patientsin total for the specific cancer. Because of the large sample size, the Z-score was reported throughout to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

provide more information for assessment of significance. For example, when compared to a Z-score of 4, a Z-score of 30 (both are $p<0.01$) may indicate a larger sample size, a larger O/E difference, or higher mortality rate of the specific cancer. Given the large sample size of the current study, the underlying assumption[13], namely the $np_0(1-p_0)>5$, was validated. We additionally evaluated the O/E ratio by categories of age, sex, race, and tumor stage; these variables were selected for their biological importance and wide application in SEER-based research. Because of the extremely low mortality (5-year RS > 99%) among stage I breast cancer and stage I-III prostate cancer, impact of the COD inaccuracy on CSS was trivial thereby the O/E ratios were not determined. Finally, the commonly reported 5-year RS was compared with the 5-year CSS to assess the impact of O/E ratio on these two net survival measures.

Results

A total of 338,445 patients were eligible for the analysis: 77,266 with breast; 95,647 with colorectal; 101,444 with lung; 29,380 with melanoma; 18,417 with prostate, and 16,291 with pancreas cancers. Baseline patient and tumor characteristics are shown in Table 1. In brief, the majority of cancer patients were aged 50+ at diagnosis, although for melanoma the median age was 53. Race was most commonly White, followed by Black and other races. Lung and pancreas cancer were more likely diagnosed with advanced than early staged disease.

The 8-year cumulative observed number of cancer-specific deaths as documented within SEER, the 8-year cumulative expected number of deaths attributed to the malignancy as estimated by the RS approach, and the corresponding O/E ratios with Z-scores for the 6 cancer sites are shown in table 2. Taking colorectal cancer as an example, the 8-year cumulative observed number of deaths that SEER documented was 60,052 (column B as shown on Table 2), whereas the 8-year cumulative expected number of overall deaths as estimated by US life tables was 19,298 (column C).The resulting difference of 40,754 (column D) was theoretically attributable to the colorectal cancer. According to the SEER COD recode variable, 39,973 deaths (column A) were actually documented as colorectal death (SEER COD recode 21040 and 21050), yielding a favorable O/E ratio of 0.98 (Z-score=5.09, $p<0.001$).Similarly, the O/E ratio for breast, lung, melanoma,prostate and pancreas cancer were 0.97, 0.90, 1.07, 1.02, and 0.92, respectively (all Z-score> 3.29, p-values< 0.001).

The O/E ratios by patient and tumor characteristics were detailed in Table 3. Because of the large sample size, the confidence interval for O/E ratios was very tight (all p-values<.05) thus data not shown. Our analyses indicated that there was little variation of the age-stratified O/E ratios. Elderly patients were generally noted to have a higher O/E ratio than younger patients, indicating that older patients were more likely to be coded as having a cancer-specific death in SEER thus their CSS were expected to be underestimated. In general, the differences of O/E ratio by sex were small, with the exception for breast cancer where the estimated O/E ratio was 0.87 (Z-score=2.33, $p<0.05$) for male and 0.97 (Z-score=3.76, $p<0.001$) for female patients, likely reflecting a COD ascertainment error given the rarity of breast cancers among men. The effect of race on the O/E ratio was also examined but the variation by race was also small. However, White race appeared to be associated with an overall favorable O/E ratio closer to 1.0 than Black race in all but prostate cancer.

For early stage cancers with favorable prognosis at baseline, the O/E ratios were more likely to be greater than 1.0, such as stage I colorectal cancer (O/E ratio=1.33, Z-score=10.9) and stage I melanoma (O/E ratio=2.12, Z-score=25.7). These findings indicate that the number

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of cancer-specific deaths as documented in SEER was over-coded by 1.33 and 2.12 times; an estimated CSS lower than RS would therefore be expected.In contrast, for cancers with generally poor prognosis (lung and pancreas cancer) or cancers in advanced stage(e.g. breast, colorectal, melanoma), SEER tended to under-coded the number of cancer-specific deaths (O/E ratio less than 1.0); an estimated CSS higher than RS would therefore be expected. The O/E ratios were also examined for patients with more than one cancer diagnosis. Not surprisingly, O/E ratios further away from 1 were consistently observed over all the 6 studied cancers. For example, patients with other cancer diagnoses in addition to the colorectal cancer had an O/E ratio of 0.84 (Z-score=23.2, $p<0.001$).

Finally, the 5-year CSS and 5-year RS were compared (Table 4). Because of the large sample size, survival estimates reported were associated with very tight confidence interval thus data not shown. Taking colorectal cancer as an example again, the results showed the 5-year CSS (58.1%) was slightly higher than the 5-year RS (57.6%). These results correspond to the fact that an O/E ratio less than 1 (0.98 for colorectal cancer) would result in an estimated CSS higher than RS. In spite of unfavorable O/E ratios for stage I colorectal cancer and melanoma as aforementioned, the differences between 5-year CSS and 5-year RS were considered small (1.9% and 1.5%, respectively). In contrast, for stage IA lung cancer, the difference (5.4%) was relatively large. For cancers with high mortality like pancreas cancer, RS and CSS were concordant with only an approximately 1% absolute difference.

Discussion

In the absence of meticulous review of primary medical records, the COD as assigned by cancer registries has long been questioned for its utility in measuring a valid CSS. In the current study of 6 common malignancies characterized by a broad range of baseline cancer-associated mortality, we quantified the COD utility by O/E ratio in overall and by categories of patient and tumor factors and accessed the O/E ratio in relation to the agreement between CSS and RS. Our study provides investigators a better understanding of the direction and extent of the bias in CSS estimation using the SEER-provided COD. For example, the calculated CSS for pancreas cancer using SEER is expected be slightly overestimated than RS but in general the difference is small.

This study employed the methodology from a prior work by Weinstock et al. The way they assessed the validity of the COD information was based on the methodology as we described earlier. They reported that, for patients diagnosed with melanoma, the COD are generally accurately certified with 4,237 expected melanoma deaths based on RS approach and 3,946 documented death according to the COD coding, representing a 93 percent concordance[10]. This methodology is less recognized but efficient in quantitatively accessing the utility of COD, which may be applicable not only to the SEER database but to other large cancer and non-cancer databases where meticulous review of medical records is not possible.

Our analysis was able to further elucidate factors that influenced the utility of the COD coding in CSS estimation. We noted that although the O/E ratio for early stage colorectal cancer was apparently poor (1.33), the resulting impact on the CSS and RS difference was trivial (95.0% vs 93.1%), because of the low underlying mortality in early stage disease. Such observation remained true for stage I melanoma, where the O/E ratio was poor (2.12) but the agreement between their RS and CSS was acceptable (98.1% vs 96.6%). Our findings suggest CSS is relatively free from the O/E disagreement for cancers with favorable prognosis. In addition, we noted that the O/E was disconcordant considerably among lung cancer. For example, the O/E ratio for stage IA lung cancer was 0.79 and the resulting difference between RS and CSS was relatively considerable (5.4%). This finding is primary

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

due to the cohort of patients at risk for lung cancer have additional tobacco exposure related non-cancer comorbidities such as cardiopulmonary disease and therefore the general population may not represent an appropriate reference population for determining expected survival[14]. As a result, CSS may be a more accurate measure of net survival than that estimated by the RS approach which fails to account for such cancer associated comorbidities.

We noted that the O/E ratio was poor for patients with multiple cancer diagnoses when compared to patients with only one cancer diagnosis. This finding emphasizes the need to exclude patients with multiple primaries in survival outcome researches using SEER. This finding was consistently observed for all 6 cancer sites and may be true for other cancer sites not under the present investigation. This finding may be related to difficulties in determining which cancer was directly attributed to the death.

The utility of the COD was subject to variations based on a patients' age at diagnosis. Specifically, we found there was a trend towards a slight overestimation of the number of documented cancer-specific death (O/E ratio>1) for elderly compared to younger patients. This is particularly noteworthy as the COD in elderly patients is often subject to speculation due to the competing effects of comorbidity-associated mortality (e.g. cardiac events or other non-cancer deaths). In sum, our finding suggested that the O/E ratio was associated with patient's age but generally the impact was small.

An alternative method to assess the COD utility examines the survival rates at a point in time at which surviving patients can be considered to have been cured of their malignancies. At such time point, the CSS and RS, as a corollary, of surviving patients approaches that of the general population. This point in time can be graphically assessed as the time when the RS or conditional survival reached a plateau[10, 15].

The use of RS as the reference for comparison assumes that RS is an unbiased outcome measure. However, an important limitation of the RS approach is the potential for non-comparability of the expected survival between the cancer groups and the matched general population[14]. For example, as mentioned earlier, for malignancies associated with certain risk factors, such as smoking and lung cancer that are highly correlated with overall health, the life table for the general population may not provide an accurate estimate of the expected survival for the lung cancer cohort. In situation like this, assessment of net survival by CSS should be advocated, unless plausible adjustment for the effect of smoking on life table could be performed.

There exist three cumulative expected methods used in calculating relative survival, including Ederer I, Ederer II and Hakulinen methods[16]. The default Ederer I method was used in the current study as this method has long been supported since late 1990s in the initial release of the SEER*Stat and has been used by the Data Analysis and Interpretation Branch of the NCI in their Cancer Statistics Review (CSR). The Hakulinen method was later supported in the early 2000 release SEER*Stat. Both methods assume matched individuals are considered to be at risk for the entire follow-up. Hakulinen additionally adjusts for potential follow-up times but, in general, relative survival estimates from these two methods are very similar[16]. In early 2011, the Ederer II method was added as the default method for relative survival estimation in SEER*Stat. This method has been revived because matched individuals are considered to be at risk only until the patient is censored or dies, thereby mitigating the potential that relative survival tends to increase in the long term when using the Ederer I and Hakulinen method[16]. Preliminary analyses using the same Figure 1 cohort showed 5-year relative survival calculated using Ederer II method (63.4%) is slightly lower than the other two methods (63.6%) but in general they are very similar for the same cohort

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Conclusions

Commonly used net survival measurements are subject to varying strengths and limitation. None of them are perfect but we found CSS as estimated based on the COD that cancer registries processed, at least for patients in whom the cancer diagnosis is their only malignancy, was generally concordant with RS. The NCI SEER*Stat program provides a convenient and intuitive mechanism to generate RS but the lack of the capability of this program to be used in regression modeling (e.g. Cox regression) and related model diagnostics is a significant limitation. Our results justify the use of CSS in survival outcomes research and the analyses can be readily performed by commonly used statistical programs.

Acknowledgments

We are grateful to the National Cancer Institute for providing the Surveillance, Epidemiology and End Results (SEER) Program.

Source of Funding: Supported by the National Institutes of Health/National Cancer Institute grants K07-CA133187 (G.J.C.) and CA016672 (MD Anderson Cancer Center's Support Grant).

List of abbreviations

COD	cause of death
CSS	cancer-specific survival
RS	relative survival
O/E ratio	observed-to-expected ratio
SEER	Surveillance Epidemiology, and End Results
NCI	National Cancer Institute

Reference

1. Surveillance, Epidemiology, and End Results (SEER) Program. [released April 2010] Research Data (1973-2007): National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch. www.seer.cancer.govbased on the November 2009 submission
2. Hoel DG, Ron E, Carter R, Mabuchi K. Influence of death certificate errors on cancer mortality trends. J Natl Cancer Inst. 1993; 85(13):1063–8. [PubMed: 8515493]
3. Smith Sehdev AE, Hutchins GM. Problems with proper completion and accuracy of the cause-of-death statement. Arch Intern Med. 2001; 161(2):277–84. [PubMed: 11176744]
4. Flanders WD. Inaccuracies of death certificate information. Epidemiology. 1992; 3(1):3–5. [PubMed: 1554807]
5. Begg CB, Schrag D. Attribution of deaths following cancer treatment. J Natl Cancer Inst. 2002; 94(14):1044–5. [PubMed: 12122090]
6. Percy C, Stanek E 3rd, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. Am J Public Health. 1981; 71(3):242–50. [PubMed: 7468855]
7. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr. 1961; 6:101–21. [PubMed: 13889176]
8. Chiang CL. A stochastic study of the life table and its applications. II. Sample variance of the observed expectation of life and other biometric functions. Hum Biol. 1960; 32:221–38. [PubMed: 13693002]
9. Gittelsohn A, Royston PN. Annotated bibliography of cause-of-death validation studies. Vital Health Stat. 1982; 2(89):1–42.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

10. Weinstock MA, Reynes JF. Validation of cause-of-death certification for outpatient cancers: the contrasting cases of melanoma and mycosis fungoides. Am J Epidemiol. 1998; 148(12):1184–6. [PubMed: 9867264]

11. Best WR, Cowper DC. The ratio of observed-to-expected mortality as a quality of care indicator in non-surgical VA patients. Med Care. 1994; 32(4):390–400. [PubMed: 8139303]

12. AJCC Cancer Staging Manual. Sixth Edition. American Joint Committee on Cancer; Springer-Verlag New York; New York: 2002.

13. Bernard, R. Fundamentals of Biostatistics. 6 ed. Duxbury Press; 2005.

14. Sarfati D, Blakely T, Pearce N. Measuring cancer survival in populations: relative survival vs cancer-specific survival. Int J Epidemiol. 39(2):598–610. [PubMed: 20142331]

15. Chang GJ, Hu CY, Eng C, Skibber JM, Rodriguez-Bigas MA. Practical application of a calculator for conditional survival in colon cancer. J Clin Oncol. 2009; 27(35):5938–43. [PubMed: 19805670]

16. Cho, H.; Howlader, N.; Mariotto, AB.; Cronin, KA. Estimating relative survival for cancer patients from the SEER Program using expected rates based on Ederer I versus Ederer II method. Surveillance Research Program, National Cancer Institute; 2011. Technical Report #2011-01Available from: http://surveillance.cancer.gov/reports/

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Condensed abstract

Cancer-specific survival is considered an acceptable surrogate for relative survival in most circumstances, justifying the use of cancer-specific survival in outcome researches which can be readily performed by commonly seen statistical programs.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

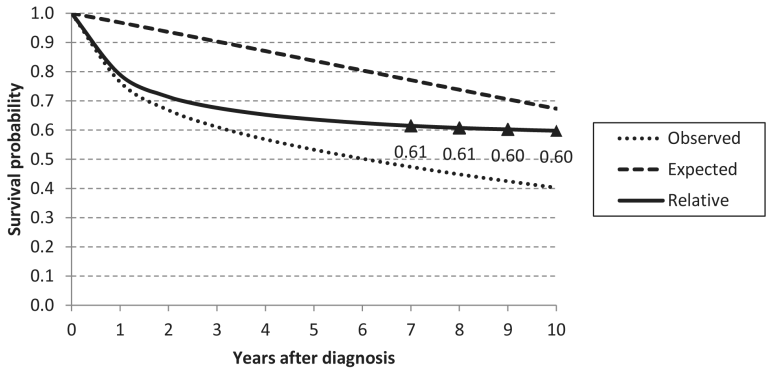

Figure 1.

Relative survival over time for patients diagnosed with breast, colorectal, lung, melanoma, prostate, and pancreas cancer between 1988 and 1997.Note the survival curve becomes flat beyond approximately 8 years after cancer diagnosis. The solid line represents the relative survival curve on which there were 61% patients surviving at year 7, 8 and 60% at years 9 and 10. There were negligible changes in RS beyond 8 years after diagnosis, indicating that the number of new deaths subsequently attributable to these cancers was negligible. Thus cancer survivors beyond 8 years of follow-up had a survival experience approximating the age, sex, race, and calendar year-matched U.S. general population. In this example using relative survival analysis, the cumulative number of deaths due to these cancers from the time of diagnosis to the time of plateau represents the number of these cancer deaths as documented within the Surveillance Epidemiology and End Results data during the same period of time, which, by our definition, would indicate an O/E ratio of 1.0.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Baseline Characteristics of the Patient Diagnosed with Breast, Colorectal, Lung, Melanoma, Prostate, and Pancreas Cancer in the 13 SEER Registries from 1988 to 1999

	Common malignancies with both low and high cancer mortality					
	Breast	**Colorectal**	**Lung**	**Melanoma***	**Prostate**	**Pancreas***
Overall, n	77,266	95,647	101,444	29,380	18,417	16,291
Age at diagnosis, **%**						
18-49	31.1	9.1	7.7	47.6	1.6	8.4
50-74	51.7	55.4	69.1	40.5	63.7	66.0
75-90	17.2	35.5	23.2	11.8	34.7	25.6
Sex, **%**						
Male	0.7	49.7	60.2	52.9	100.0	51.1
Female	99.3	50.3	39.8	47.1	-	48.9
Race, **%**						
White	81.9	82.6	81.0	95.1	77.6	80.3
Black	10.3	9.0	11.9	0.6	15.5	12.1
Other	7.8	8.4	7.1	4.3	6.9	7.6
Tumor Stage, **%**						
I	-	16.5	-	78.0	-	6.1
IA	-	-	8.1	-	-	-
IB	-	-	9.9	-	-	-
II	-	-	-	13.3	-	23.5
IIA	49.9	26.0	1.3	-	-	-
IIB	26.3	5.6	5.4	-	-	-
III	-	-	-	4.2	-	11.4
IIIA	7.6	2.6	8.6	-	-	-
IIIB	7.2	16.0	25.2	-	-	-
IIIC	-	9.6	-	-	-	-
IV	9.0	23.8	41.4	4.5	100.0	59.0

*Tumor stages were regrouped into I, II, III, IV as some substages were sparse

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Detailed Calculation of Observed-to-Expected Ratio

Cancer sites	Column A	Column B	Column C	Column D	O/E ratio (column A ÷ column D)	Z-score*
	Eight-year cumulative observed number of cancer-specific deaths documented within SEER	Eight -year cumulative observed number of overall deaths that SEER documented	Eight -year cumulative expected number of overall deaths as estimated by US life tables	Eight-year cumulative expected number of deaths attributed to the malignancy (column B - column C)		
Breast (stage II-IV)	23,216	33,133	9,409	23,724	0.97	3.95
Colorectal	39,973	60,052	19,298	40,754	0.98	5.09
Lung	79,535	94,182	6,732	87,450	0.90	70.1
Melanoma	4,045	7,197	3,446	3,751	1.07	5.12
Prostate (stage IV)	10,092	19,198	9,400	9,798	1.02	4.29
Pancreas	14,322	16,436	945	15,491	0.92	39.1

Abbreviations: SEER, Surveillance Epidemiology and End Results

*Z-score> 1.96, p-value< 0.05; Z-score> 2.58, p-value< 0.01; Z-score> 3.29, p-value< 0.001

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Cancer. Author manuscript; available in PMC 2014 May 15.

Table 3

Observed-to-Expected Ratio, by Patient and Tumor Characteristics

Characteristics	Breast		Colorectal		Lung		Melanoma		Prostate		Pancreas	
	O/E ratio	Z-score[a]	O/E ratio	Z-score	O/E ratio	Z-score	O/E ratio	Z-score	O/E ratio	Z-score	O/E ratio	Z-score
Overall	0.97	3.95	0.98	5.09	0.90	70.1	1.07	5.12	1.02	4.29	0.92	39.1
Age (years)												
18-49	0.96	4.12	0.92	6.11	0.89	24.3	0.97	1.04	0.95	1.09	0.91	14.1
50-74	0.97	3.44	0.96	8.08	0.90	65.5	1.19	8.59	1.00	0.71	0.91	37.9
75-90	1.02	2.10	1.03	4.83	0.93	20.2	1.00	0.26	1.07	6.56	0.94	12.0
Sex												
Male	0.86	2.33	0.97	4.42	0.91	56.3	1.07	4.12	1.02	4.29	0.92	28.3
Female	0.97	3.76	0.98	2.76	0.90	42.6	1.08	3.11	-	-	0.92	27.0
Race												
White	0.99	0.99	0.99	1.52	0.91	59.9	1.07	4.56	1.04	6.06	0.93	31.5
Black	0.92	5.78	0.95	4.45	0.90	28.7	0.80	1.88	0.99	0.10	0.89	18.4
Other	0.92	3.87	0.89	8.06	0.88	24.2	2.83	9.47	0.89	3.75	0.87	15.9
AJCC 6th Tumor Stage												
I	-	-	1.33	10.9	-	-	2.12	25.7	-	-	0.95	2.85
IA	-	-	-	-	0.79	17.8	-	-	-	-	-	-
IB	-	-	-	-	0.87	15.7	-	-	-	-	-	-
II	-	-	-	-	-	-	0.95	2.20	-	-	0.95	9.54
IIA	1.09	7.58	1.04	3.59	0.88	6.35	-	-	-	-	-	-
IIB	0.96	3.31	0.96	1.81	0.88	16.7	-	-	-	-	-	-
III							0.96	1.53	-	-	0.94	10.8
IIIA	0.96	2.39	1.21	5.05	0.91	20.4	-	-	-	-	-	-
IIIB	0.93	6.33	1.00	0.86	0.92	41.7	-	-	-	-	-	-
IIIC	-	-	0.97	3.44	-	-	-	-	-	-	-	-
IV	0.92	16.2	0.93	29.0	0.91	79.3	0.83	15.2	1.02	4.29	0.90	46.3
Number of primaries												

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Characteristics	Breast		Colorectal		Lung		Melanoma		Prostate		Pancreas	
	O/E ratio	Z-score[a]	O/E ratio	Z-score	O/E ratio	Z-score	O/E ratio	Z-score	O/E ratio	Z-score	O/E ratio	Z-score
First and only	0.97	3.95	0.98	5.09	0.90	70.1	1.07	5.12	1.02	4.29	0.92	39.1
With multiple primaries (exclude first and only)	0.83	16.5	0.84	23.2	0.81	49.1	0.83	6.83	0.67	17.9	0.86	21.8

Abbreviations: AJCC, American Joint Committee on Cancer

[a]Z-score> 1.96, p-value< 0.05; Z-score> 2.58, p-value< 0.01; Z-score> 3.29, p-value< 0.001

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

Comparison of Five-year Relative Survival and Disease-Specific Survival

Characteristics	Breast		Colorectum		Lung		Melanoma		Prostate		Pancreas	
	RS	CSS	RS	CSS	RS	CSS	RS	CSS	RS	CSS	RS	CSS
Overall	74.2	74.7	57.6	58.1	13.3	16.1	88.4	88.0	49.3	48.2	2.8	3.7
Age (years)												
18-49	75.6	76.4	57.5	59.4	13.8	16.8	91.6	92.0	43.5	45.1	5.5	6.9
50-74	74.1	74.9	57.4	58.7	13.8	16.9	86.5	85.3	52.8	52.8	2.6	3.5
75-90	70.8	70.9	58.0	56.9	10.9	13.2	78.4	79.7	40.4	38.7	2.0	2.9
Sex												
Male	68.7	73.0	55.6	56.2	11.3	14.1	85.0	84.6	49.3	48.2	2.6	3.4
Female	74.2	74.7	59.5	59.9	16.1	19.0	92.1	91.8	-	-	3.0	3.9
Race												
White	75.6	75.9	58.3	58.5	13.6	16.4	88.1	87.8	50.5	48.9	2.7	3.4
Black	60.3	62.1	48.1	49.7	10.2	13.0	67.0	70.8	41.3	41.4	2.0	3.3
Other	77.1	78.7	60.5	62.8	13.9	17.0	97.2	94.6	53.1	55.1	4.8	6.6
AJCC 6th Tumor Stage												
I	-	-	95.0	93.1	-	-	98.1	96.6	-	-	14.9	15.5
IA	-	-	-	-	58.6	64.0	-	-	-	-	-	-
IB	-	-	-	-	42.3	45.8	-	-	-	-	-	-
II	-	-	-	-	-	-	68.6	70.2	-	-	5.6	6.1
IIA	89.8	89.1	83.0	81.8	33.7	38.0	-	-	-	-	-	-
IIB	76.2	76.7	68.3	68.5	22.1	26.1	-	-	-	-	-	-
III	-	-	-	-	-	-	46.9	47.7	-	-	1.7	2.4
IIIA	63.1	63.9	83.5	81.2	11.3	13.6	-	-	-	-	-	-
IIIB	36.5	39.2	59.9	59.3	4.7	6.1	-	-	-	-	-	-
IIIC			40.0	40.4			-	-	-	-	-	-
IV	17.0	19.2	5.4	6.3	1.3	2.0	10.9	15.8	49.3	48.2	0.6	1.3
Multiple primaries												
First and only	74.2	74.7	57.6	58.1	13.3	16.1	88.4	88.0	49.3	48.2	2.8	3.7

Characteristics	Breast		Colorectum		Lung		Melanoma		Prostate		Pancreas	
	RS	CSS	RS	CSS	RS	CSS	RS	CSS	RS	CSS	RS	CSS
With multiple primaries (excluded first and only)	78.4	81.5	68.8	72.2	32.3	39.2	90.0	91.3	59.7	69.4	7.4	11.6

Abbreviations: AJCC, American Joint Committee on Cancer; CSS, disease-specific survival; RS, relative survival.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript