Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jonathan Barkley Date

Using Conditional Survival to Examine Poor-Prognosis Cancers in the U.S.

By

Jonathan Barkley

Master of Public Health

Epidemiology

_____

Kevin C. Ward, PhD, MPH

Thesis Committee Chair

Using Conditional Survival to Examine Poor-Prognosis Cancers in the U.S.

By

Jonathan Barkley

Bachelor of Arts

St. Mary's College of Maryland

2012

Thesis Committee Chair: Kevin C. Ward, PhD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2014

Abstract

Using Conditional Survival to Examine Poor-Prognosis Cancers in the U.S.

By Jonathan Barkley

Background: Conditional survival estimates show that patients diagnosed with poor-prognosis cancer types have a much improved outlook if able to survive the first few years of high mortality. The primary aim of this investigation was to better understand the clinical and demographic characteristics that influence survival and how these associations change as patients survive past diagnosis.

Methods: Data were obtained from the Surveillance, Epidemiology, and End Results (SEER) program November 2012 submission. Five-year cause-specific survival conditional on surviving 0-5 years was estimated for 13 cancers using the period method in SEER*Stat. The five-year hazard of death was modeled at diagnosis (overall model) and 2 years past diagnosis (conditional model) for pancreatic and esophageal cancers diagnosed between 2001 and 2009 using Cox proportional hazard (PH) models. Age, sex, race/ethnicity, marital status, grade, stage, region, urban-rural status, and treatment were controlled for in each model. After exclusions for missing data, the final overall and conditional models for esophageal cancer consisted of 23,383 and 7,592 patients respectively, while models for pancreatic cancer consisted of 63,380 and 10,140 patients respectively.

Results: Five-year conditional survival increased by the greatest magnitude for pancreatic (+67%) and esophageal cancers (+61%), 2 of the cancers with the lowest 5-year survival at diagnosis. Cox PH models identified increasing age, increasing stage, and the absence of treatment as statistically associated with an increased hazard of death in both the overall and conditional cohorts. For esophageal cancer, marital status, race, and grade were associated with hazard of death in the overall model, but were no longer significant in the conditional model. For pancreatic cancer, a significant interaction between marital status and race was observed for the conditional model, where patients of other races who were married had a significantly lower hazard compared to whites (HR=0.73, 95% CI: 0.60, 0.88) versus those who were non-married (HR=0.96, 95% CI: 0.96, 1.67).

Conclusions: The risk profile of patients diagnosed with esophageal and pancreatic cancer changes as these patients survive past diagnosis. Prognostic models such as these could be used to obtain individualized hazard estimates and could assist clinicians in patient counseling during the survivorship period.

Using Conditional Survival to Examine Poor Prognosis Cancers in the U.S.

By

Jonathan Barkley

Bachelor of Arts

St. Mary's College of Maryland

2012

Thesis Committee Chair: Kevin C. Ward, PhD, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2014

Acknowledgements

I would like to thank Dr. Kevin Ward for the guidance, helpful consults, and feedback that he provided throughout the entire process of completing this project. Both your knowledge and time were very much appreciated.

Table of contents

Background

In the United States, cancer is the number one cause of death among people under the age of 85, where males have a one in two lifetime risk and females have a one in three lifetime risk of developing the disease (1). Due to this large burden, estimating and understanding site-specific cancer survival is critical so that clinicians can effectively communicate with their patients regarding prognosis and research efforts can be targeted to better understand disparities in outcomes that might exist. In general, cancer survival is measured from the time one is diagnosed with the disease until some specified period in the future, typically 5 years for most cancers, and can be estimated using crude or net measures.

Crude measures of survival attempt to estimate the probability of survival in the presence of all causes of death while net measures focus on survival in the absence of other causes of death. Net survival thus attempts to control for competing causes of mortality and more accurately reflects the experiences of those diagnosed with cancer. Two types of net survival, relative survival and cause-specific survival control for competing mortalities in different ways. Relative survival compares the survival experience of a defined disease cohort to that of a matched group from the general population to obtain a ratio of observed to expected survival rates. This is a popular measure typically used by cancer registries because it does not rely on death certificates for a coded cause of death and can avoid potential misclassification. For example, cause of death could be inaccurately attributed to the site of metastasis on a death certificate instead of the underlying primary site where the cancer originated. However, suitable general population life tables that are necessary to calculate relative survival are not

always available for all groups, particularly racial/ethnic minorities or different socioeconomic classes. As such, cause-specific survival measures are also widely used (2).

Cause-specific survival only considers mortality from a specific cause (i.e., death from stomach cancer for stomach cancer patients) and thus does rely on knowing the cause of death and assuming it has been accurately coded. Individuals who die from other causes are considered to be censored in the analyses. In recent years, the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute has developed new methods to improve the accuracy of cause-specific survival estimates. SEER uses two different cause-specific death classification variables that independently identify deaths dues to cancer and deaths due to other causes. These variables are defined using an algorithm that, in addition to death certificate coded cause of death, considers tumor sequence, site of original cancer diagnosis, and comorbidities in order to capture deaths that were thought to be related to the cancer of interest but were not coded in such a way (3).

Several approaches can be used to derive the net survival estimates discussed above. These include the cohort, multiple year, and period methods. Each of these approaches is described below and illustrated in Figure 1 (4). The cohort method utilizes the survival experience of patients who were diagnosed within the same calendar year, or group of calendar years, and have complete follow-up for the entire cohort over time. This method has been seen to produce out of date survival estimates because patients who are more recently diagnosed with a cancer often do not have the same survival experience as this comparison cohort which was diagnosed several years prior. Multiple year and

period survival estimates overcome this limitation by using the most recently available information on a cohort of cancer patients to derive estimates of survival. This is important as it appropriately reflects current patterns of diagnosis, treatment and cancer patient care. Specifically, the multiple year method (aka the complete method), allows patients diagnosed in different calendar years to be included in the same cohort, where each individual can contribute varying lengths of survival time up until the study cutoff of interest. This means, for example, that patients diagnosed in the past 2 years could still contribute to 5-year survival estimates even though they have not been followed for a full 5 years. In contrast to this method, period survival uses only the most recent interval of data that is available for patients diagnosed within a specific calendar year, or group of calendar years. For example, to obtain a 5-year survival estimate for a particular cancer in the year 2009, those diagnosed in 2008 would contribute to the 1 year probability of survival; those diagnosed in 2007 would contribute to the 2 year probability of survival, and so forth. Five-year period survival would then be calculated by multiplying the survival probabilities for each of the 5 intervals, as shown below in Figure 1 (4). In the figure, patients diagnosed with regional breast cancer in 2001 contribute data for the 1-year survival interval (96.8%); those diagnosed in 2000 contribute data for the 2-year survival interval (94.4%); those diagnosed in 1999 contribute data for the 3-year survival interval (94.7%); those diagnosed in 1998 contribute data for the 4-year survival interval (94.0%), and those diagnosed in 1997 contribute data for the 5-year survival interval (94.1%). The product of these 5 corresponding proportions produces the 5-year net survival of 76.6% that is displayed in the figure. As can be seen, this method provides the most up-to date estimate of survival available by allowing the left-most observations

to be truncated, while right-censoring the observations at the end. Nationally representative studies investigating survival from several cancer types in Japan, Australia, and Canada have utilized this approach. (5-7). These studies will be discussed in greater detail.

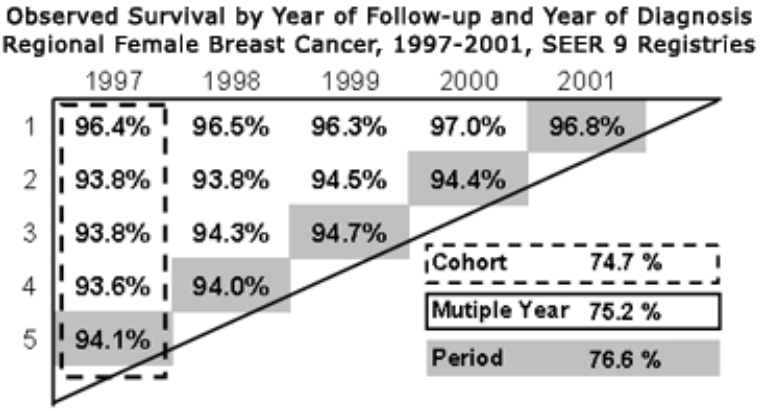


Figure 1. Illustrations of the different ways to obtain 5-year survival estimates from patients diagnosed with a cancer between 1997 and 2001. Adapted from the SEER website. Available at: <http://surveillance.cancer.gov/survival/cohort.html>.

Traditional survival methods provide an overall picture of the average survival experience of patients newly diagnosed with a specific cancer. For example, a recent study in Canada estimated the 5-year survival of patients diagnosed with stomach cancer to be 24% (6). Traditional estimates are less informative, however, for people who survive 1 or more years past their diagnosis, as the risk of cancer death is typically greatest during the first few years. While traditional methods measure outcomes beginning from the time of diagnosis, conditional methods measure survival for those who have already survived a given time period. Conditional survival is thus defined as the probability of living an additional number of years given the patient has survived a

specified number of years past diagnosis. It is a dynamic method that has the added value of better understanding how a patient's survival experience changes over time. For example, the conditional 5-year survival of stomach cancer in the example referenced above was 92% among the subset of patients who survived 5 years past their initial diagnosis (6). While it is acknowledged that the size of the cohort will decrease over time due to mortality, conditional estimates can be used to help shape the outlook of patients who survive the early years of their diagnosis. Conditional survival can be applied to any of the survival methodologies described earlier. Evidence from the literature suggests that specific benefits of using conditional survival estimates include helping to better understand the changing risk profile of patients, playing a role during patient counseling, and assisting in the comparison of rates between studies and countries (8). These benefits will be discussed below.

Understanding a patient's changing risk profile may be clinically important in developing treatment and counseling strategies. Mortality from fatal cancers is highest early on, because these cancers are usually diagnosed at later stages. However, patients who survive through the early years of higher mortality often have a more favorable outlook moving forward. For these patients, it is important that this change in risk be communicated. Mortality after the first year or two is often due to late effects and comorbid conditions, and influences due to these factors can be captured using conditional survival (8). Specifically, lower conditional survival suggests patients may suffer from late mortality from advanced stage disease or late treatments effects. The role that late effects play in each cancer type may influence future treatment options and follow-up periods suggested by clinicians. For example, for good prognosis cancer types

with few late effects, patients who survive 5 years past diagnosis now have similar survival rates as cancer-free members of the population (8). Thus, follow-up periods for these patients could be less frequent. However, for cancer types with fair and poor prognosis, conditional survival can help clinicians understand the time intervals where late effects often play a role and more appropriately schedule surveillance follow-up accordingly.

Estimates of conditional survival can also be used during patient counseling and to help patients make life decisions that are not related to treatment, as they continue to live past their diagnosis (9). For example, if patients understand the manner in which survival for most cancers improves conditionally over time, they may be less prone to anxiety and better able to plan for the future understanding that they are no longer at the same risk compared to when they were diagnosed. In fact, web-based tools have been developed for some cancers (rectal and head and neck cancers) where patients can enter their personal characteristics and current time past diagnosis in order to obtain updated future survival estimates (9, 10). It is unclear the extent to which such tools are being utilized and similar tools do not appear to exist for other cancer types. The utility of conditional survival estimates for both clinicians and patients suggests they should be more widely used (11).

It has also been described that absolute differences in survival estimates can become narrower when using conditional survival compared to overall survival estimates (12) . For example, when comparing colorectal cancer patient 5- year survival between various countries , there was a 12.3% difference in the 5-year overall survival, but only a 6% difference when comparing the corresponding five-year conditional survival

estimates (12). The smaller degree of variation in conditional estimates makes these estimates attractive for comparison purposes. Since healthcare systems and treatment strategies differ between countries, differences in conditional estimates may help better understand treatment effectiveness, specifically for cancers that typically have poor prognoses. As mentioned earlier, late effects due to treatment or comorbidities can be the cause of declining conditional survival. Thus, by comparing conditional survival estimates between countries, researchers may better understand the relative influences of factors that affect survival.

In the literature, conditional survival estimates are typically calculated one through five years after diagnosis and generally focus on a single cancer. As noted by Shack and colleagues, five years is a good upper limit for this calculation as it is able to capture the majority of the improvement seen in most cancers (8). However, some studies have looked at longer-term survival to see whether conditional survival estimates change (13, 14). Results from these studies support that five-year conditional survival is generally sufficient. Regardless of the cancer of interest, most studies consistently show conditional survival improves over time and that the greatest improvements are seen for poor-prognosis cancers or for advanced stages of disease. For example, five-year survival of resected pancreatic adenocarcinoma was only 21%, but the five-year survival conditional on having survived five years reaches 71% (15). Other cancers investigated using conditional survival methodology include the following: Ewing's sarcoma (16), colon cancer (17), brain cancer (18, 19), head and neck cancer (14, 20), melanoma (21, 22), ovarian cancer (23), gastric cancer (24), and pancreatic duct cancer (25).

Fewer studies, on the other hand, have looked at conditional survival across multiple cancer sites, and most of these studies have occurred overseas. Studies conducted in Europe, the United States, Canada, Australia, and Japan have found similar results, although identical cancer types were not included in all studies (5-7, 26, 27). Specifically, the study in the United States looked at 11 cancer sites that were diagnosed between 1990 and 2001 and followed through 2006 (27). Between these studies, results were typically stratified by age, sex, and stage (5-7, 26). As expected, the greatest improvements in conditional survival were generally seen for in cancers with poor prognosis and with advanced stage at diagnosis. Some exceptions were observed. For example, chronic lymphocytic leukemia (CLL) was included in the Canadian study and was the only cancer not to exhibit improved conditional survival (6). This lack of improvement is likely due to etiology specific to this disease. Conditional survival percentages typically were above 90% after 5 years, but lower conditional survival was observed for some of the fair and poor-prognosis cancers. These lower estimates suggest that factors that influence late mortality play a greater role in these cancer types.

Most of the studies in the literature to date have not used multivariate modeling to predict conditional survival. Of the international studies described earlier, only Merrill and colleagues controlled for covariates using multivariate modeling. In this study, separate regression models were constructed for each cancer stage and site, controlling for age, sex, race, year of diagnosis, and years since diagnosis (27). All of the other national studies reported conditional survival stratified by combinations of age, gender, and stage (5-7, 26). Four studies used multivariate Cox proportional hazard models to predict the hazard of the death at different conditional intervals, however all of these

studies focused on a single cancer type (19, 21, 22, 25). Additionally, only two of these studies were conducted on cancer types with poorer prognosis (19, 25). Covariates that were generally included in the models included stage, age, year of diagnosis, ethnicity, sex, and residence. Marital status was considered in two of these studies (21, 25). These covariates may serve as a basis for variables that may want to be controlled for in future investigations that look at multiple cancer sites.

Age, ethnicity, and other factors have been discussed in other studies investigating conditional relative survival. Conditional survival improvement of young adults was found to be less than older adults (28). In this study, cancers of young adults were compared to cancers of older adults and young adults between age 20 and 29 consistently had lower conditional survival estimates. However, some evidence suggests that the effect of covariates may diminish over time. For example, Janssen-Heijnen and colleagues looked at long term survival in the Netherlands and found that although initial differences were found due to age, gender, and stage, these differences disappeared after surviving between five and ten years (14). The role of race/ethnicity in conditional survival is similar to that observed in traditional survival studies. One study looked at racial/ethnic disparities in lung cancer and found that blacks tended to have lower conditional survival rates than whites (24). These disparities persisted even five years after diagnosis suggesting race/ethnicity or other unmeasured factors associated with race/ethnicity continue to drive some of the observed survival disparities using both traditional and conditional measures.

In conclusion, evidence from the literature suggests that conditional survival estimates are of clinical importance and offer an added benefit to traditional survival

estimates alone. However, not many national studies investigating multiple cancer types have been conducted using these measures, and few studies have used multivariate modeling techniques to support their results. Thus, it is of interest to investigate conditional survival using the most recently available SEER data, to compare these estimates to previous studies in the United States and other countries, and to use mathematical modeling to see how conditional survival estimates are modified by important covariates.

Objectives

The primary aims of this research are 1) to estimate current cancer survival probabilities in the United States using recent SEER data, and 2) to construct survival models for select cancers to better understand the influence important covariates play in net survival versus net conditional survival. Meeting these objectives will allow for a better understanding of differences that exist between overall and conditional survival, and provide insight into which covariates may explain observed differences.

Methods

Estimating survival of a variety of cancers

The first aim of this project was to explore current estimates of cancer survival using SEER data. Cause-specific survival for 5 years, and for 5 years conditional on surviving 1-5 years, was estimated using data from the 18 SEER cancer registries available through SEER*Stat (version 8.1.2) (29). The period method of survival analysis was used to obtain the most up-to-date survival estimates for 18 cancers, which were representative of diseases with good, medium, and poor prognoses. Period survival estimates were generated using Kaplan-Meier tables, where cause of death was defined as death due to cancer using SEER's cause-specific death classification (3). Cases were selected into the cohort if they were diagnosed between 1998 and 2009, actively followed, had malignant behavior, were of known age, and were in the research database. First primary only cases were selected. Cases with missing or unknown cause of death were excluded from analysis. Cases were further excluded from the cohort if they were identified only by death certificate or autopsy, or were alive with no survival time. In order to ensure sufficient cases were accumulated, each period interval was selected to contain 3 years of data. The study cutoff date for survival calculations was set as December 31, 2010. Special intervals were entered in SEER*Stat in order to obtain the 1-5 year conditional survival estimates, in additional to the overall 5-year estimates.

Modeling overview

Survival of select poor-prognosis cancers was modeled to better elucidate the effects important covariates play in conditional versus traditional survival. Two Cox proportional hazard models were constructed for each cancer using SAS 9.3 (Cary, NC).

One model was constructed to predict the 5-year hazard of death at diagnosis and another was constructed to predict the 5-year hazard of death, conditional on having survived 2 years. Modeling survival using the period method was computationally challenging, so the multiple year method of survival was used instead. As described previously, this method allows individuals that have contributed survival time to be included in the cohort, although they may not have been followed for 5 years. Thus, this method allows for more up-to-date estimates than the cohort method and can be modeled using Cox regression. The cohort for the models constructed comprised individuals diagnosed with the particular cancer of interest from 2001 through 2009, as this range of years allows for 5 years of survival data to be tracked for part of the cohort, even after conditioning on surviving 2 years past diagnosis. Thus, the first model (overall survival model) contained cases diagnosed from 2001 and 2009, and the second model (conditional survival model) contained individuals from the first model who survived at least 2 years post-diagnosis.

The first cancer modeled was esophageal cancer. Using SEER*Stat, esophageal cancer cases diagnosed between 2001 and 2009 with a first primary cancer were exported into SAS. Similar to the first analysis, cases were excluded from the cohort if they were identified only by death certificate or autopsy, or were alive with no survival time. The following SEER variables were included in the dataset: sex, age at diagnosis, race, Hispanic ethnicity, marital status at diagnosis, SEER registry, urban-rural status, cancer sequence number, stage, grade, primary site, and treatment.

SEER variables were re-coded into the following classifications: age at diagnosis (0-55, 55-70, 70+), marital status (married, other), region (West, Midwest, Northeast, Southeast), urban-rural status (metropolitan, urban, rural), stage (local, regional, distant,

unknown), grade (1&2, 3&4, unknown), primary site (cervical/upper, thoracic/middle, abdominal/lower, overlapping/not otherwise specified), treatment (no treatment versus some type of treatment, including both radiation and surgery).

Regional classifications were defined as follows: West (Seattle-Puget Sound, San-Francisco-Oakland, San Jose-Monterey, Los Angeles, Alaska natives, and Hawaii), Midwest (Utah and New Mexico), Northeast (Iowa, Detroit, Connecticut, New Jersey), and Southeast (Kentucky, Atlanta, Georgia, rural Georgia, and Louisiana). Urban-rural status was assigned using the urban code classifications available on the SEER website (30). The decision was made to exclude individuals with unknown/missing marital status, race, and urban status as these classifications represented small percentages of the total within each variable. Individuals with unknown treatment data were included in the no treatment group.

A 'status' variable was created to distinguish death events (death attributed to esophageal cancer) versus censored events (alive at end of follow-up, untraced, death due to cause besides esophageal cancer). For the unconditional model, status was coded as 1 if a patient survived less than or equal to 60 months after diagnosis and died from esophageal cancer (or common metastatic sites, as classified by SEER's cause-specific death variable). Otherwise, status was coded as 0. The unconditional model included patients who had survived at least 2 years past diagnosis. Similarly for this model, status was coded as 1 if a patient survived less than or equal to 84 months after diagnosis and died from esophageal cancer. Otherwise, status was coded as 0.

The second cancer modeled was pancreatic cancer. Variables were coded identically as described for esophageal cancer with the exception of the primary site variable, which was coded to account for the following groups: head of pancreas, body of pancreas, tail of pancreas, other/overlapping/unknown areas of the pancreas.

In the overall hazard model for esophageal cancer, 985 patients (4.0%) with unknown marital status, 98 patients (0.4%) with unknown race information, and 43 patients (0.18%) with unknown urban-rural status were excluded. In the conditional survival model for esophageal cancer, 325 patients (4.1%) with unknown marital status , 41 patients (0.5%) with unknown race information, and 13 patients (0.16%) with unknown urban-rural status were excluded. In the overall model for pancreatic cancer, 2,159 patients (3.4%) with unknown marital status, 192 patients (0.3%) with unknown race information, and 86 patients (0.13%) with unknown urban-rural status were excluded. In the conditional model for pancreatic cancer, 377 patients (3.56%) with unknown marital status, 64 patients (0.60%) with unknown race, and 9 patients with missing urban-rural status were excluded. After exclusions, the final cohort used to create the overall and conditional hazard models for esophageal cancer consisted of 23,383 and 7,592 patients, respectively. The final cohort used to create the overall and conditional survival models for pancreatic cancer consisted of 63,380 and 10,140 patients, respectively.

For both esophageal and pancreatic cancer, univariate differences between covariates were tested between those who died from the cancer versus those who were censored for each data set using Chi-square tests in SAS (Cary, NC). Cox proportional hazard models were constructed using SAS to model the 5-year hazard of death for the

overall and conditional datasets for each cancer. Before creating the model, each variable was examined using log-log survival curves to test for violations of the proportional hazards assumption. Next, interaction between selected covariates of interest (i.e. marital status with sex, race, and age) was assessed using the backward elimination approach, where likelihood ratio tests were performed comparing reduced and full models. No-interaction Cox proportional hazard models were run for both the overall and conditional esophageal datasets. Due to evidence of interaction in both pancreatic cancer datasets, interaction terms were added to each respective model and stratified results were presented where appropriate.

Results

Five-year survival conditional on having survived 0 through 5 years was first estimated for several cancers using the period method (Table 1). Estimates are presented in ascending order by traditional 5-year survival measures with cancers of poorest prognosis listed first. As an example of what can be seen from these data, it is observed in Table 1 that while 20.4% of esophageal cancer patients were estimated to survive 5 years past diagnosis (0 year conditional), 82% of the patients were estimated to survive another 5 years, conditional on having survived the first 5 years (5 year conditional). In general, conditional survival improved across all cancer types with the largest improvements in 5-year survival observed for cancers with poor prognosis. Going from a conditional survival of 0 to 5 years, the greatest improvements in 5-year survival were observed for pancreatic cancer (+67.0%), followed by esophageal cancer (+61.6%). These were the 1st and 3rd most deadly cancers respectively at diagnosis. The smallest improvements in conditional survival tended to be among cancers of good prognosis. Specifically, the lowest improvements in 5-year survival were observed in prostate cancer (+0.9%), followed by thyroid cancer (+1.3%).

Next, 2 cohorts were developed for both esophageal and pancreatic cancer patients. Those diagnosed with each cancer were included in the 1st cohort (overall cohort), and those who were still alive 2 years after diagnosis were included in the 2nd cohort (conditional cohort). Characteristics of each cohort were compared by vital status (dead vs. alive/censored) at 5 years using univariate comparisons (Tables 2 and 3). For esophageal cancer measuring survival from diagnosis, significant differences existed by race ($p<0.0001$, Table 2a). Specifically among those who were deceased, blacks

comprised a larger proportion (13.0% vs. 10.0%) while whites comprised a smaller proportion (82.1% vs. 85.5%) relative to their alive/censored counterparts. However, among patients who survived 2 years past diagnosis, significant racial differences no longer existed ($p=0.470$, Table 2b). In the overall cohort, deceased patients were less likely to be diagnosed in the Northeast ($p<0.0001$), more likely to have been diagnosed at age 70 or older (42.2% vs. 36.9%, $p<0.0001$) and more likely to not have received treatment (83.1% vs. 52.0%, $p<0.0001$). However, no significant differences by region ($p=0.330$), age ($p=0.095$) or treatment ($p=0.054$) remained after conditioning on survival for 2 years. As it relates to marital status, deceased patients in the overall cohort were less likely to be married compared to their alive counterparts ($p<0.0001$), while deceased patients in the conditional cohort were more likely to be married than their alive counterparts ($p=0.021$). In both the overall and conditional models, patients who were deceased were more likely to have been diagnosed with higher grades and stages of cancer ($p<0.0001$). No significant differences by vital status were observed for sex, Hispanic ethnicity, or urban-rural status among patients from either cohort (Tables 2a & 2b).

With regards to pancreatic cancer, deceased patients in the overall cohort did not differ from those who were censored with respect to sex, but differed across all other covariates (Table 3a). Specifically, those who died were more likely to be white, Non-Hispanic, greater than 70 years of age, diagnosed in West or Midwest regions, and diagnosed with higher grades and stages of cancer than their censored counterparts. All of these differences were also observed among those who had survived at least 2 years past diagnosis (Table 3b). Patients in the overall cohort who died were less likely to have

lived in metropolitan areas than their censored counterparts, while the converse was true for patients living in urban and rural areas ($p=0.0002$, Table 3a). However, conditional on 2 year survival, vital status no longer differed across metropolitan, urban, or rural areas ($p=0.211$, Table 3b). As it relates to marital status, deceased patients were less likely to be married in the overall cohort ($p<0.0001$, Table 3a), but more likely to be married in the conditional cohort (Table 3b).

Next, Cox proportional hazard models were constructed to examine the relationship of individual covariates on the hazard of death while controlling for other factors of interest and to explore how these relationships changed over time. Log-log survival curves showed that no variable grossly violated the proportional hazards assumption for either esophageal or pancreatic cancer. Interaction assessment found no significant interaction in either model for esophageal cancer, thus reduced models are reported (Tables 4a & 4b). For pancreatic cancer, the backwards elimination approach suggested evidence for interaction between marital status and sex for the overall 5-year hazard model at diagnosis, and evidence for interaction between marital status and race for the 5-year hazard model that was conditional on surviving 2 years past diagnosis. Thus, effects due to these interactions were controlled for when obtaining the parameter estimates for each model (Tables 5a and 5b). The individual effects of the interaction terms are presented separately in Table 6, controlling for all other factors in the model. Several findings were consistent between both the overall and conditional survival models for each of the cancers. For example, patients diagnosed with regional or distant stage cancers had significantly higher hazards of death compared to patients diagnosed at local stages in all models (Tables 4 & 5). The strength of this effect was larger in the

conditional models. In addition, patients who underwent some form of treatment (radiation or surgery vs. no treatment) had a significantly lower hazard of death, although the effect was less strong in the conditional model, while patients diagnosed with cancer over the age of 55 had significantly higher hazards compared to patients diagnosed before age 55.

In the overall 5–year hazard model for esophageal cancer, patients who were not married had a significantly higher hazard of death compared to those who were married (Table 4a, HR=1.21, 95% CI: 1.17, 1.25). Additionally, blacks had a significantly higher hazard of death than whites (HR=1.08, 95% CI: 1.03,1.14), patients living in rural areas had a significantly higher hazard of death that those living in metropolitan areas (HR=1.18, 95% CI: 1.05,1.33), and patients with advanced grades of disease (3 or 4) had a significantly higher hazard of death than those with lower grades of disease (1 or 2) (HR=1.25, 95% CI: 1.21,1.29). However, all these differences became not significant once the cohort was conditioned on having survived 2 years (Table 4b).

In the overall hazard model for pancreatic cancer, patients living in urban areas had a significantly higher hazard of death compared to those living in metropolitan areas (Table 5a), however this difference became non-significant once the cohort was conditioned on having survived 2 years (Table 5b). In contrast to esophageal cancer, increased mortality due to higher grades of disease persisted in both models ($p<0.0001$, Tables 5a & 5b). Interaction effects for the pancreatic cancer hazard models are presented separately in Table 6. For the overall hazard model, females had lower hazards of death compared to males across both stratums of marital status, however non-married females tended to have slightly lower hazards of death (HR=0.90, 95% CI=0.87, 0.92) than

married females (HR=0.93, 95% CI= 0.91, 0.95). For the conditional hazard model, there was no difference in the hazards between married blacks and whites or non-married blacks and whites ($p>0.05$, Table 6). However, married patients of other races had a significantly lower hazard of death compared to whites (HR=0.73, 95% CI: 0.60, 0.88), while non-married patients of other races did not show the same effect (HR=1.26, 95% CI: 0.96, 1.67). For esophageal cancer, other races had lower, but non-significantly different hazards compared to white across both stratums of marital status (data not presented).

Discussion

Conditional survival estimates show that patients diagnosed with poor-prognosis cancer types have a much-improved outlook if able to survive the early years of high mortality. Although a smaller cohort of patients is present after this time, from a risk profile perspective, it is important to focus on these individuals and better understand the characteristics of their cohort. A primary aim of this paper was to investigate the changing role clinical and demographic variables play once patients have survived past their initial diagnosis. The overall 5-year hazard of death at diagnosis and 5-year hazard of death after surviving 2 years were modeled using Cox proportional hazard models for 2 poor prognosis cancers (esophageal and pancreatic).

Before modeling was conducted, up-to date 5-year net survival estimates were obtained for several cancers using the period method, conditional on surviving 0-5 years past diagnosis. Consistent with previous studies in the U.S. and Canada (6, 27), the greatest gains in conditional survival were observed for cancer types with poor prognosis. Pancreatic cancer showed the greatest magnitude of increase in both of these studies, consistent with our results showing a 67% increase in conditional survival after 5 years compared to survival at diagnosis. The Canadian study noted that survival from cancer types with poor prognosis typically remained below 90% (6). Although the study used relative survival to obtain survival estimates compared to our study, which used cause-specific survival estimates, consistent results were observed. Minor differences have been found to exist between these national-level studies. For example, a study in Japan found conditional survival from liver cancer to increase only modestly (13%), while our study observed a much larger magnitude of improvement (54.4%). Such differences may be

attributed to different types of liver cancer, different treatment strategies between countries, delayed diagnosis, or influences due to comorbidities or late effects. Nevertheless, improvements in conditional survival tended to be greatest in all the national-level studies among patients who were diagnosed with poor prognosis cancers (5, 6, 26, 27), supporting the belief that traditional survival estimates become less relevant for these cancers as patients continue to be survive past their diagnosis. Thus, conditional measures could be useful in clinical settings. Specifically, if physicians reported these measures to patients during follow-up visits past initial diagnosis, patients could better understand their changing risk profile and adjust their views for the future.

The high conditional survival of these patients relative to the overall survival of the cohort suggests it is important to study the factors that influence survival in these individuals. However, few studies have modeled the effect covariates play in survival of poor prognosis cancers, and these studies have generally relied on multivariate regression techniques. The few studies that utilized Cox proportional hazard models have focused solely on one cancer type and considered a limited number of covariates (7, 19, 22, 25). Thus, in this study we aimed to better elucidate the role several key demographic and clinical variables may play in the survival of two generally poor prognosis cancer types: esophageal and pancreatic cancer. The results from our models are consistent with current knowledge of these cancers, but also present some interesting implications to consider when analyzing future cohorts and trying to more widely implement these approaches for counseling programs.

Initial univariate analyses for each of these cancers by vital status suggested that race, region, and age at diagnosis may play a role among survival of patients at diagnosis.

Specifically, for esophageal cancer, vital status differed by all of these factors at diagnosis, but no longer differed conditional on 2 year survival. Among those deceased, patients in the overall cohort for esophageal cancer and the conditional cohort for pancreatic cancer were more likely to be from the West region. Interestingly for both esophageal and pancreatic cancer, the deceased were more likely to be non-married at diagnosis relative to their alive counterparts, but more likely to be married conditional on surviving 2 years. Vital status differed significantly by some characteristics, such stage and treatment, in both the overall and conditional cohorts. These observations suggest these covariates may influence survival so we further explored these associations were through our Cox proportional hazard models.

The Cox models confirmed several of the findings from our univariate analyses. Consistent with the univariate results for esophageal cancer, blacks had a significantly higher hazard compared to whites at diagnosis, but not 2 years after diagnosis suggesting this race effect is strongest early on. Effects of region were stronger at diagnosis for both cancers, suggesting that regional differences in health systems could help influence early survival from these poor prognosis cancers. The Northeast region in particular sticks out as the hazard of death was significantly lower for both models in esophageal cancer and in the overall model for pancreatic cancer. Similar to the univariate analyses, marital status was associated with a significantly lower hazard of death at diagnosis. However, any negative effect of marital status that was suggested by the univariate analyses was not present once we controlled for the covariates in our models.

As expected, older age (70+), distant stages of disease, and receiving no treatment were associated with statistically significantly higher hazards of death among

esophageal and pancreatic cancer patients. These differences were observed in both the overall and conditional Cox proportional hazard models. In the three models that included sex, males had a significantly higher hazard of death compared to females. All of these results have been confirmed in previous studies (6, 27). For esophageal cancer, several factors associated with an increased hazard of death at diagnosis were no longer significantly associated with the hazard of death once patients survived 2 years past diagnosis. These factors included race, region, urban-rural status, grade, and marital status. Many of these effects are also consistent with other studies which found initial differences to disappear after surviving several years past diagnosis (6, 13, 19, 26). Specifically, unmarried versus married, black versus white race/ethnicity, rural versus metropolitan area of residence, and higher grades of disease were associated with an increased hazard of death in the overall model, but were no longer significant in the conditional model. Factors such as these that are significant early on, but not past diagnosis, may suggest they play a role in the quality of patient care or patient support networks, but not in the overall disease course (19).

The role of marital status has only been considered in a few studies modeling conditional hazards, and these studies generally have focused on a single cancer type (19, 21, 22, 25). However, a previous study found that the survival benefit from marriage exceeded the benefit for chemotherapy for esophageal and other cancer types (31). These findings suggest marriage or other forms of social support may play a critical role by influencing a patient to seek treatment for an illness or have improved outlooks. One study that examined conditional Cox proportional hazard models among patients with glioblastoma found marital status was associated with the hazard of death at diagnosis,

but no longer remained a significant predictor 1 and 3 years after diagnosis (19). These findings are consistent with our observations for esophageal cancer.

The observed interaction between marital status and race in the conditional Cox proportional hazard model for pancreatic cancer provides further insight into the role marriage may play in this population. Interestingly, married patients of other races had a significantly lower hazard of death than married whites after surviving 2 years past diagnosis. However, no significant differences were present between non-married whites and non-married patients of other races. These results suggest married patients of other races gain the greatest survival advantage compared to whites, thus the effect of marriage as social support is very important in this group. Similar associations by marital status were not observed for esophageal cancer (data not presented), thus potential interaction effects between marital status and race are important to consider in future studies that utilize Cox proportional hazard models.

There are a few limitations of our study that exist mainly due to general issues that persist when working with registry-based data. The potential for misclassification of cancers exists, however any effect would likely be minor as almost all cases are microscopically confirmed. Misclassification could also be introduced due to errors in SEER's cause-specific survival death classification variable, as it is possible that cause of death is not always accurately coded. We were also not able to consider the effects of comorbidities, socioeconomic status, or treatments beyond surgery and radiation therapy, which would have been informative and may have introduced some bias into our estimates. General limitations due to the data coding that were used also exist. For example, in order to obtain sufficient sample sizes, categories such as grades I and II

were grouped together into a single category along with all surgery and radiation treatment options. Although such categorizations could introduce some residual confounding we believe any effects would likely be small since groupings were made based on the knowledge of disease etiology.

However, there are several advantages to using the methodological approach we used. The greatest advantage of SEER registry data is that large cohorts of patients can be considered. This is especially helpful when investigating poor prognosis cancers since cases are relatively rare in the general population. The SEER program is also representative of the U.S. The program currently covers 28% of the U.S. population, and overall the SEER population is similar to the general population with respect to race and socioeconomic status (32) Recent improvements have also been made in the SEER cause-specific death classification variable which make misclassification less likely. For example, SEER uses two different variables to reach this classification and accounts for death at common metastatic sites for each cancer type (3). Furthermore, by using the period method to estimate conditional survival, and considering cases diagnosed between 2001 and 2009 when creating the cohorts for Cox proportional hazard modeling, we are able to obtain up-to-date estimates for cancer survival/mortality. This investigation also has the advantage of being able to compare associations and patterns across more than one cancer type.

In this study we found factors such as age, race, and marital status play an important role in early survival from poor prognosis cancer types. The finding related to marital status have some interesting implications Marital status likely serves as a proxy for social support, thus placing an emphasis on patient support programs may be one way

that initial survival could be improved, irrespective of advances in treatment. Since initial survival differences due to race are present, it is also worthy to investigate health and treatment disparities to see if progress can be made.

Significance and Conclusions

The main significance of this investigation is that it supports the value of using conditional models in clinical settings. Our results for both esophageal and pancreatic cancer consistently show that the survival experience for patients who live a year or two past their initial diagnosis is much improved compared to experience of the overall population at diagnosis, and that different variables are important predictors for the hazard of death depending on where a patient falls on this continuum. Thus, it is important to use these methodologies to communicate this improved outlook to patients and better understand their changing risk profiles. By understanding which factors influence survival of poor prognosis cancers, prognostic models similar to the ones presented in this paper can be constructed to obtain individualized survival estimates. The greatest benefit of these models likely would be if used by clinicians to communicate to patients how their survival experience improves as they move through the continuum of cancer survival. By better understanding patient's hazards, clinicians can more effectively communicate risk and potentially use these models to help schedule follow-up visits.

The use of conditional models could also play a role in counseling patients as they continue to live past their initial diagnosis and have to make life decisions. The fear of recurrence is a major cause of anxiety of cancer survivors along with their caregivers (33, 34). Furthermore, the stress of cancer patients has been shown to impact their overall quality of life (35). If patients better understand how their survival experience changes and improves past diagnosis then they could have a more positive outlook toward the future. In fact, web-based tools have been developed for a couple of cancers (rectal and

head and neck cancers) where patients can enter their personal characteristics and current time past diagnosis in order to obtain updated future survival estimates (9, 10). If similar tools were more widely implemented and available for additional cancer types, they may be able to play a greater role in patient counseling and decision making. The utility of this methodology in both clinical and counseling settings stresses the importance of reporting conditional survival probabilities and modeling hazard using conditional models. As more cancers are investigated and factors associated with short-term and longer-term survival are validated across cancers perhaps similar models will become more relevant and more widely used by clinicians.

In conclusion, although most patients diagnosed with poor prognosis cancer types do not survive several years after their diagnosis, it is important to study the patients who do survive past the period of high mortality. By understanding the demographic and clinical composition of these patients we can better learn what influences survival and how patients risk profiles change over time. As seen in this investigation, some factors such as marital status may have the largest impact on early survival, while other factors such as stage and treatment continue to influence mortality even two years past diagnosis. As a patient moves through the continuum of cancer survival these models could be useful to both clinicians and patients as they aim to better communicate and understand factors that influence survival.

References

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin 2014;64(1):9-29.

2. Howlader N, Ries LA, Mariotto AB, Reichman ME, Ruhl J, Cronin KA. Improved estimates of cancer-specific survival rates from population-based data. J Natl Cancer Inst 2010;102(20):1584-98.

3. National Cancer Institute. SEER cause-specific death classification. http://seer.cancer.gov/causespecific/. Accessed April 12, 2014.

4. National Cancer Institute. Cancer survival statistics: cohort definition using diagnosis year. http://surveillance.cancer.gov/survival/cohort.html. Updated December 21, 2010. Accessed April 12, 204.

5. Ito Y, Nakayama T, Miyashiro I, Ioka A, Tsukuma H. Conditional survival for longer-term survivors from 2000-2004 using population-based cancer registry data in Osaka, Japan. BMC Cancer 2013;13:304.

6. Ellison LF, Bryant H, Lockwood G, Shack L. Conditional survival analyses across cancer sites. Health Rep 2011;22(2):21-5.

7. Yu XQ, Baade PD, O'Connell DL. Conditional survival of cancer patients: an Australian perspective. BMC Cancer 2012;12:460.

8. Shack L, Bryant H, Lockwood G, Ellison LF. Conditional relative survival: a different perspective to measuring cancer outcomes. Cancer Epidemiol 2013;37(4):446-8.

9. Bowles TL, Hu CY, You NY, Skibber JM, Rodriguez-Bigas MA, Chang GJ. An individualized conditional survival calculator for patients with rectal cancer. Dis Colon Rectum 2013;56(5):551-9.

10. Wang SJ, Wissel AR, Ord CB, Kalpathy-Cramer J, Fuller CD, Holland JM, et al. Individualized estimation of conditional survival for patients with head and neck cancer. Otolaryngol Head Neck Surg 2011;145(1):71-3.

11. Zabor EC, Gonen M, Chapman PB, Panageas KS. Dynamic prognostication using conditional survival estimates. Cancer 2013;119(20):3589-92.

12. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. Lancet 2011;377(9760):127-38.

13. Janssen-Heijnen ML, Houterman S, Lemmens VE, Brenner H, Steyerberg EW, Coebergh JW. Prognosis for long-term survivors of cancer. Ann Oncol 2007;18(8):1408-13.

14. van der Schroeff MP, van de Schans SA, Piccirillo JF, Langeveld TP, Baatenburg de Jong RJ, Janssen-Heijnen ML. Conditional relative survival in head and neck squamous cell carcinoma: Permanent excess mortality risk for long-term survivors. Head Neck 2010;32(12):1613-8.

15. Mishra MV, Champ CE, Keith SW, Showalter TN, Anne PR, Lawrence YR, et al. Conditional Survival Probabilities for Patients With Resected Pancreatic Adenocarcinoma. Am J Clin Oncol 2012.

16. Miller BJ, Lynch CF, Buckwalter JA. Conditional survival is greater than overall survival at diagnosis in patients with osteosarcoma and Ewing's sarcoma. Clin Orthop Relat Res 2013;471(11):3398-404.

17. van Steenbergen LN, Steur M, Lemmens VE, Rutten HJ, van Spronsen DJ, Janssen-Heijnen ML. Minimal excess mortality for long-term colon cancer survivors in the Netherlands 1989-2008. Eur J Cancer 2013;49(3):585-92.

18. Porter KR, McCarthy BJ, Berbaum ML, Davis FG. Conditional survival of all primary brain tumor patients by age, behavior, and histology. Neuroepidemiology 2011;36(4):230-9.

19. Johnson DR, Ma DJ, Buckner JC, Hammack JE. Conditional probability of long-term survival in glioblastoma: a population-based analysis. Cancer 2012;118(22):5608-13.

20. Fuller CD, Wang SJ, Thomas CR, Jr., Hoffman HT, Weber RS, Rosenthal DI. Conditional survival in head and neck squamous cell carcinoma: results from the SEER dataset 1973-1998. Cancer 2007;109(7):1331-43.

21. Xing Y, Chang GJ, Hu CY, Askew RL, Ross MI, Gershenwald JE, et al. Conditional survival estimates improve over time for patients with advanced melanoma: results from a population-based analysis. Cancer 2010;116(9):2234-41.

22. Bowles TL, Xing Y, Hu CY, Mungovan KS, Askew RL, Chang GJ, et al. Conditional survival estimates improve over 5 years for melanoma survivors with node-positive disease. Ann Surg Oncol 2010;17(8):2015-23.

23. Choi M, Fuller CD, Thomas CR, Jr., Wang SJ. Conditional survival in ovarian cancer: results from the SEER dataset 1988-2001. Gynecol Oncol 2008;109(2):203-9.

24. Wang SJ, Fuller CD, Thomas CR, Jr. Ethnic disparities in conditional survival of patients with non-small cell lung cancer. J Thorac Oncol 2007;2(3):180-90.

25. Gleason MX, Mdzinarishvili T, Are C, Sasson A, Sherman A, Shats O, et al. Prognostic estimator of survival for patients with localized and extended pancreatic ductal adenocarcinoma. Cancer Inform 2013;12:103-14.

26. Janssen-Heijnen ML, Gondos A, Bray F, Hakulinen T, Brewster DH, Brenner H, et al. Clinical relevance of conditional survival of cancer patients in europe: age-specific analyses of 13 cancers. J Clin Oncol 2010;28(15):2520-8.

27. Merrill RM, Hunter BD. Conditional survival among cancer patients in the United States. Oncologist 2010;15(8):873-82.

28. Bleyer A, Choi M, Fuller CD, Thomas CR, Jr., Wang SJ. Relative lack of conditional survival improvement in young adults with cancer. Semin Oncol 2009;36(5):460-7.

29. Surveillance, Epidemiology, and End Results (SEER) Program, SEER*Stat Database: Incidence - SEER 18 Regs Research. November 2012 submission. httpp://www.seer.cancer.gov. Accessed March 2014.

30. National Cancer Institute. Rural-urban continuum codes. http://seer.cancer.gov/seerstat/variables/countyattribs/ruralurban.html. Updated April 11, 2008. Accessed March 11, 2014.

31. Aizer AA, Chen MH, McCarthy EP, Mendu ML, Koo S, Wilhite TJ, et al. Marital status and survival in patients with cancer. J Clin Oncol 2013;31(31):3869-76.

32. National Cancer Institute. Overview of the SEER program. http://seer.cancer.gov/about/overview.html. Accessed April 16, 2014.

33. Mellon S, Kershaw TS, Northouse LL, Freeman-Gibb L. A family-based model to predict fear of recurrence for cancer survivors and their caregivers. Psychooncology 2007;16(3):214-23.

34. Mullens AB, McCaul KD, Erickson SC, Sandgren AK. Coping after cancer: risk perceptions, worry, and health behaviors among colorectal cancer survivors. Psychooncology 2004;13(6):367-76.

35. Ferrell BR, Dow KH, Grant M. Measurement of the quality of life in cancer survivors. Qual Life Res 1995;4(6):523-31.

Tables

Table 1. 5-year cause-specific survival estimates for select cancers conditional on having survived 0-5 years [1]. Estimates obtained using the period method where each interval included 3 years of data. Cases diagnosed between 1998-2011.

	Survival (%) conditional on surviving (years)					
Cancer type	***0***	***1***	***2***	***3***	***4***	***5***
Pancreas	7.0	22.5	41.1	55.5	66.9	74.0
Lung and Bronchus	19.6	39.0	53.5	62.9	69.1	73.7
Esophageal	20.4	39.1	57.9	69.3	76.6	82.0
Liver	22.4	44.3	56.3	64.8	70.8	76.8
Stomach	30.9	53.2	68.4	77.1	81.6	85.6
Brain	32.8	52.3	68.3	76.5	80.8	83.1
Ovary	45.1	55.2	61.0	66.5	72.5	77.3
Colon	64.9	75.2	80.4	84.6	87.7	90.2
Rectum	68.3	74.4	78.5	82.5	85.9	88.2
Non-Hodgkin's Lymphoma	71.1	83.4	86.6	88.0	88.5	89.2
Kidney and Renal	74.9	84.7	87.7	89.2	90.1	90.5
Urinary Bladder	78.5	85.5	89.3	90.9	91.9	92.5
Corpus Uteri	83.0	87.9	91.4	93.6	95.3	96.4
Hodgkin's Lymphoma	87.6	92.5	94.0	95.0	95.5	96.2
Breast (female)	88.0	89.0	89.9	90.9	91.7	92.5
Melanoma	89.8	91.8	93.5	94.8	95.5	96.0
Prostate	93.9	94.4	94.7	94.8	94.9	94.8
Thyroid	97.0	98.3	98.4	98.5	98.5	98.3

1. Source: SEER 18 registry data + hurricane Katrina impacted Louisiana cases, November 2012 submission (1973-2010 varying) - Linked to county attributes - Total U.S., 1969-2011 counties.

Table 2a. Characteristics of those who died from esophageal cancer or were censored [1] among individuals diagnosed with the disease from 2001-2009

Characteristic	Status=0[a] n	(%)	Status=1[b] n	(%)	χ^2 p-value
Sex					0.683
Male	4,900	(77.5)	14,007	(77.2)	
Female	1,424	(22.5)	4,129	(22.8)	
Marital-status					<0.0001
Married	3,733	(61.8)	9,613	(55.1)	
Other	2,307	(38.2)	7,822	(43.2)	
Race					<0.0001
White	5,375	(85.5)	14,837	(82.1)	
Black	626	(10.0)	2,347	(13.0)	
Other	286	(4.6)	891	(4.9)	
Hispanic Ethnicity					0.683
Hispanic	437	(6.9)	1,226	(6.8)	
Non-Hispanic	5,887	(93.1)	16,910	(93.2)	
Age at diagnosis (y)					<0.0001
0-54	1,667	(18.5)	3,185	(17.6)	
55-70	2,822	(44.6)	7,298	(40.2)	
70+	2,823	(36.9)	7,653	(42.2)	
Region					<0.0001
West	2,650	(41.9)	7,971	(44.0)	
Midwest	226	(3.6)	717	(4.0)	
Northeast	2,060	(32.6)	5,039	(27.8)	
Southeast	1,388	(22.0)	4,409	(24.3)	
Urban-rural status					0.110
Metropolitan	5,479	(86.8)	15,596	(86.1)	
Urban	751	(11.9)	2,209	(12.2)	
Rural	82	(1.3)	300	(1.7)	
Grade					<0.0001
1 and 2	2,681	(42.4)	6,067	(33.5)	
3 and 4	2,207	(34.9)	8,437	(46.5)	
Unknown	1,436	(22.7)	3,632	(20.0)	
Stage					
Local	2,499	(39.5)	2,692	(14.8)	<0.0001
Regional	2,154	(34.1)	5,142	(28.4)	
Distant	936	(14.8)	7,882	(43.5)	
Unknown	735	(11.6)	2,420	(13.3)	
Primary site					<0.0001
Cervical/upper esophagus	422	(6.7)	1,244	(6.9)	
Thoracic/middle esophagus	1,100	(17.4)	3,602	(19.9)	
Abdominal/lower esophagus	3,979	(62.9)	10,340	(57.0)	
Overlapping/unknown region	823	(13.0)	2,950	(16.3)	
Treatment					<0.0001
No treatment	3,290	(52.0)	15,074	(83.1)	
Treatment	3,034	(48.0)	3,062	(16.9)	

1. At 5 years
a. Censored individuals (includes those who were alive, untraced, or died from cause other than esophageal cancer).
b. Death due to esophageal cancer.

Table 2b. Characteristics of those who died from esophageal cancer or were censored [1] among individuals diagnosed with the disease from 2001-2009, given they had survived at least 2 years.

Characteristic	Status=0[a] n	(%)	Status=1[b] n	(%)	χ^2 p-value
Sex					0.232
Male	4,808	(77.5)	1,375	(78.8)	
Female	1,400	(22.6)	370	(21.2)	
Marital-status					0.021
Married	3,660	(61.7)	1,102	(64.8)	
Other	2,268	(38.3)	598	(35.2)	
Race					0.470
White	5,270	(85.4)	1,506	(86.5)	
Black	616	(10.0)	164	(9.4)	
Other	285	(4.6)	71	(4.1)	
Hispanic Ethnicity					0.202
Hispanic	431	(6.9)	106	(6.1)	
Non-Hispanic	5,777	(93.1)	1,639	(93.9)	
Age at diagnosis (y)					0.095
0-54	1,154	(18.6)	320	(18.3)	
55-70	2,760	(44.5)	824	(47.2)	
70+	2,294	(37.0)	601	(34.4)	
Region					0.330
West	2,604	(42.0)	754	(43.2)	
Midwest	222	(3.6)	75	(4.3)	
Northeast	2014	(32.4)	548	(31.4)	
Southeast	1,368	(22.0)	368	(21.1)	
Urban-rural status					0.505
Metropolitan	5,380	(86.8)	1,523	(87.3)	
Urban	736	(11.9)	194	(11.1)	
Rural	80	(1.3)	27	(1.6)	
Grade					0.0003
1 and 2	2,626	(42.3)	734	(42.1)	
3 and 4	2,164	(34.9)	681	(39.0)	
Unknown	1,418	(22.8)	330	(18.9)	
Stage					<0.0001
Local	2,463	(39.7)	453	(26.0)	
Regional	2,102	(33.9)	755	(43.3)	
Distant	927	(14.9)	348	(19.9)	
Unknown	716	(11.5)	189	(10.8)	
Primary site					0.070
Cervical/upper esophagus	421	(6.8)	102	(5.9)	
Thoracic/middle esophagus	1072	(17.3)	321	(18.4)	
Abdominal/lower esophagus	3,907	(62.9)	1,127	(64.6)	
Overlapping/unknown region	808	(13.0)	195	(11.2)	
Treatment					0.054
No treatment	3,239	(52.2)	956	(54.8)	
Treatment	2,969	(47.8)	789	(45.2)	

1. At 5 years
a. Censored individuals (includes those who were alive, untraced, or died from cause other than esophageal cancer).
b. Death due to esophageal cancer.

Table 3a. Characteristics of those who died from pancreatic cancer or were censored [1] among individuals diagnosed with the disease from 2001-2009.

Characteristic	Status=0[a] n	(%)	Status=1[b] n	(%)	χ^2 p-value
Sex					0.108
Male	3,836	(48.2)	28,463	(49.2)	
Female	4,117	(51.8)	29,368	(50.8)	
Marital-status					<0.0001
Married	4,391	(57.4)	30,308	(54.2)	
Other	3,254	(42.6)	25,644	(45.8)	
Race					<0.0001
White	6,176	(78.2)	46,839	(81.2)	
Black	1,061	(13.4)	6,904	(12.0)	
Other	657	(8.3)	3,928	(6.8)	
Hispanic Ethnicity					<0.0001
Hispanic	884	11.1	5,306	(9.2)	
Non-Hispanic	7,069	(88.9)	52,498	(90.8)	
Age at diagnosis					<0.0001
0-54	1,790	(22.5)	7,641	(13.2)	
55-70	2,826	(35.5)	19,518	(33.8)	
70+	3,337	(42.0)	30,645	(53.0)	
Region					<0.0001
West	3,584	(45.1)	27,295	(42.2)	
Midwest	324	(4.1)	2,415	(4.2)	
Northeast	2,441	(30.7)	15,922	(27.5)	
Southeast	1,604	(20.2)	12,172	(21.1)	
Urban-rural status					0.0002
Metropolitan	7,098	(89.3)	50,655	(87.8)	
Urban	760	(9.6)	6,227	(10.8)	
Rural	89	(1.12)	842	(1.5)	
Grade					<0.0001
1 and 2	2,557	(32.2)	8,730	(15.1)	
3 and 4	1,022	(12.9)	8,581	(14.8)	
Unknown	4,374	(55.0)	40,483	(70.0)	
Stage					<0.0001
Local	1,654	(20.8)	3,776	(6.5)	
Regional	2,951	(37.1)	14,629	(25.3)	
Distant	2,454	(30.9)	32,404	(56.1)	
Unknown	894	(11.2)	6,995	(12.1)	
Primary site					<0.0001
Head	4,231	(53.2)	27,599	(47.8)	
Body	648	(8.2)	5,712	(9.9)	
Tail	1,131	(14.2)	6,404	(11.1)	
Other/Overlapping/ Unknown	1,943	(24.4)	18,089	(31.3)	
Treatment					<0.0001
No treatment	4,421	(55.6)	51,053	(88.3)	
Treatment	3,532	(44.4)	6,751	(11.7)	

1 At 5 years
a Censored individuals (includes those who were alive, untraced, or died from cause other than pancreatic cancer).
b Death due to pancreatic cancer.

Table 3b. Characteristics of those who died from pancreatic cancer or were censored [1] among individuals diagnosed with the disease from 2001-2009, given they had survived at least 2 years .

Characteristic	Status=0[a] n	(%)	Status=1[b] n	(%)	χ^2 p-value
Sex					0.009
Male	3,751	(48.3)	1,438	(51.2)	
Female	4,018	(51.7)	1,372	(48.8)	
Marital-status					<0.0001
Married	4,270	(57.2)	1,829	(67.0)	
Other	3,200	(42.8)	903	(33.1)	
Race					<0.0001
White	6,026	(78.2)	2,342	(83.5)	
Black	1,036	(13.4)	272	(9.7)	
Other	648	(8.4)	191	(6.8)	
Hispanic Ethnicity					0.0008
Hispanic	865	(11.1)	249	(8.9)	
Non-Hispanic	6,904	(88.9)	2,561	(91.1)	
Age at diagnosis					<0.0001
0-54	1,742	(22.4)	600	(21.4)	
55-70	2,750	(35.4)	1,275	(45.4)	
70+	3,277	(42.2)	935	(33.3)	
Region					<0.0001
West	3,502	(45.1)	1,350	(48.0)	
Midwest	320	(4.1)	92	(3.3)	
Northeast	2,380	(30.6)	903	(32.1)	
Southeast	1,567	(20.2)	465	(16.6)	
Urban-rural status					0.211
Metropolitan	6,927	(89.2)	2,538	(90.4)	
Urban	747	(9.6)	241	(8.6)	
Rural	89	(1.2)	28	(1.0)	
Grade					<0.0001
1 and 2	2,468	(31.8)	1,076	(38.3)	
3 and 4	1,005	(12.9)	452	(16.1)	
Unknown	4,296	(55.3)	1,282	(45.6)	
Stage					<0.0001
Local	1,629	(21.0)	295	(10.5)	
Regional	2,866	(36.9)	1,489	(53.0)	
Distant	2,401	(30.9)	770	(27.4)	
Unknown	873	(11.2)	256	(9.1)	
Primary site					<0.0001
Head	4,124	(53.1)	1,707	(60.8)	
Body	638	(8.2)	241	(8.6)	
Tail	1,102	(14.2)	283	(10.1)	
Other/Overlapping/Un known	1,905	(24.5)	579	(23.3)	
Treatment					<0.0001
No treatment	4,356	(56.1)	1,437	(51.1)	
Treatment	3,413	(43.9)	1,373	(48.9)	

1 At 5 years
a Censored individuals (includes those who were alive, untraced, or died from cause other than pancreatic cancer).
b Death due to pancreatic cancer.

Table 4a. Esophageal cancer hazard ratio estimates from overall 5-year survival model. Model included 23,383 cases diagnosed from 2001 to 2009.

Variable	HR (95% CI)	p-value
Sex		
Male	1.00	-
Female	0.93 (0.89,0.96)	0.0001
Marital-status		
Married	1.00	-
Other	1.21 (1.17.1.25)	<.0001
Race		
White	1.00	-
Black	1.08 (1.03,1.14)	0.001
Other	0.94 (0.88,1.01)	0.119
Hispanic Ethnicity		
Non-Hispanic	1.00	-
Hispanic	0.99 (0.93,1.05)	0.810
Age at diagnosis (y)		
0-54	1.00	-
55-70	1.04 (1.00,1.09)	0.053
70+	1.36 (1.30,1.42)	<.0001
Region		
West	1.00	-
Midwest	1.04 (0.96,1.12)	0.398
Northeast	0.89 (0.86,0.92)	<.0001
Southeast	1.05 (1.01,1.10)	0.020
Urban-rural status		
Metropolitan	1.00	-
Urban	1.01 (0.96,1.05)	0.841
Rural	1.18 (1.05,1.33)	0.006
Grade		
1 and 2	1.00	-
3 and 4	1.25 (1.21,1.29)	<.0001
Unknown	1.02 (0.98,1.07)	0.298
Stage		
Local	1.00	-
Regional	1.57 (1.49.1.64)	<.0001
Distant	2.63 (2.51,2.76)	<.0001
Unknown	1.75 (1.65,1.85)	<.0001
Primary site		
Cervical/upper esophagus	1.00	-
Thoracic/middle esophagus	1.09 (1.02,1.16)	0.013
Abdominal/lower esophagus	1.04 (0.98,1.11)	0.210
Overlapping/unknown region	1.24 (1.16,1.33)	<.0001
Treatment		
No treatment	1.00	-
Treatment	0.43 (0.41,0.44)	<.0001

Table 4b. Esophageal cancer hazard ratio estimates from 5-year conditional survival model. Model included the 7,592 cases from the initial cohort that had survived at least 2 years.

Variable	HR (95% CI)	p-value
Sex		
Male	1.00	-
Female	0.84 (0.74,0.95)	0.005
Marital-status		
Married	1.00	-
Other	1.07 (0.97,1.19)	0.187
Race		
White	1.00	-
Black	0.98 (0.82,1.17)	0.839
Other	0.75 (0.58,0.96)	0.021
Hispanic Ethnicity		
Non-Hispanic	1.00	-
Hispanic	0.89 (0.72,1.09)	0.251
Age at diagnosis (y)		
0-54	1.00	-
55-70	1.24 (1.08,1.41)	0.002
70+	1.47 (1.27,1.70)	<.0001
Region		
West	1.00	-
Midwest	1.11 (0.87,1.42)	0.420
Northeast	0.88 (0.79,0.99)	0.029
Southeast	1.02 (0.89,1.17)	0.775
Urban-rural status		
Metropolitan	1.00	-
Urban	0.88 (0.75,1.03)	0.101
Rural	1.32 (0.90,1.95)	0.156
Grade		
1 and 2	1.00	-
3 and 4	1.07 (0.96,1.19)	0.203
Unknown	0.91 (0.79,1.04)	0.159
Stage		
Local	1.00	-
Regional	1.98 (1.75,2.23)	<.0001
Distant	3.01 (2.59.3.50)	<.0001
Unknown	1.73 (1.44,2.08)	<.0001
Primary site		
Cervical/upper esophagus	1.00	-
Thoracic/middle esophagus	1.28 (1.02, 1.60)	0.034
Abdominal/lower esophagus	1.39 (1.13,1.72)	0.002
Overlapping/unknown region	1.28 (1.00,1.64)	0.052
Treatment		
No treatment	1.00	-
Treatment	0.61 (0.55,0.68)	<.0001

Table 5a. Pancreatic cancer hazard ratio estimates from overall 5-year survival model. Model included 63,380 cases diagnosed from 2001 to 2009. Parameter estimates of terms involved in interaction are not displayed.

Variable	HR (95% CI)	p-value
Age		
0-54	1.00	-
55-70	1.29 (1.25,1.32)	<.0001
70+	1.79 (1.75,1.84)	<.0001
Race		
White	1.00	-
Black	1.04 (1.01,1.07)	0.004
Other	0.96 (0.93,0.99)	0.017
Hispanic ethnicity		
Non-Hispanic	1.00	-
Hispanic	1.00 (0.97,1.03)	0.815
Region		
West	1.00	-
Midwest	1.06 (1.01,1.10)	0.011
Northeast	0.92 (0.91,0.94)	<.0001
Southeast	1.06 (1.03,1.08)	<.0001
Urban-rural status		
Metropolitan	1.00	-
Urban	1.05 (1.02,1.08)	0.001
Rural	1.07 (0.99,1.14)	0.079
Grade		
1 and 2	1.00	-
3 and 4	1.42 (1.38,1.46)	<.0001
Unknown	1.19 (1.16,1.22)	<.0001
Stage		
Local	1.00	-
Regional	1.34 (1.29,1.39)	<.0001
Distant	2.09 (2.02,2.16)	<.0001
Unknown	1.46 (1.40,1.53)	<.0001
Primary site		
Head	1.00	-
Body	0.98 (0.95,1,01)	0.104
Tail	1.01 (0.98,1.04)	0.414
Other/Overlapping/ Unknown	1.10 (1.08, 1.12)	<.0001
Treatment		
No treatment	1.00	-
Treatment	0.41 (0.40,0.42)	<.0001

Table 5b. Pancreatic cancer hazard ratio estimates from 5-year conditional survival model. Model included the 10,140 cases from the initial cohort that had survived at least 2 years. Parameter estimates of terms invovled in interaction are not displayed.

Variable	HR (95% CI)	p-value
Sex		
Male	1.00	-
Female	0.91 (0.84,0.98)	0.018
Age		
0-54	1.00	-
55-70	1.53 (1.39,1.70)	<.0001
70+	1.56 (1.40,1.73)	<.0001
Hispanic ethnicity		
Non-Hispanic	1.00	-
Hispanic	0.85 (0.74,0.98)	0.022
Region		
West	1.00	-
Midwest	0.80 (0.65,1.00)	0.051
Northeast	1.00 (0.91,1.09)	0.936
Southeast	0.87 (0.78,0.98)	0.023
Urban-rural status		
Metropolitan	1.00	-
Urban	0.94 (0.81,1.08)	0.372
Rural	1.17 (0.80,1.70)	0.421
Grade		
1 and 2	1.00	-
3 and 4	1.25 (1.12,1.40)	<.0001
Unknown	0.85 (0.77,0.93)	0.001
Stage		
Local	1.00	-
Regional	2.50 (2.20,2.85)	<.0001
Distant	2.34 (2.02,2.70)	<.0001
Unknown	1.66 (1.39,1.99)	<.0001
Primary site		
Head	1.00	-
Body	1.06 (0.92,1.22)	0.429
Tail	0.70 (0.61,0.80)	<.0001
Other/Overlapping/ Unknown	0.87 (0.78,0.96)	0.006
Treatment		
No treatment	1.00	-
Treatment	0.51 (0.46,0.56)	<.0001

Table 6.. Pancreatic cancer hazard ratio estimates statified by marital status for terms that were involved in interaction in either the overall model (marital status and sex) or conditional model (marital status and race). 63,380 cases diagnosed between 2001 and 2009 included in the overall model, and 10,140 cases included in the conditional model.

		Married			***Not married***	
Model	**Variable**	**HR (95% CI)**	**p-value**	**Variable**	**HR (95% CI)**	**p-value**
Overall	*Sex* [a]			*Sex*		
	Male	1.00	-	Male	1.00	-
	Female	0.93 (0.91,0.95)	<.0001	Female	0.90 (0.87,0.92)	<.0001
Conditional	*Race* [b]			*Race*		
	White	1.00	-	White	1.00	-
	Black	0.90 (0.75,1.10)	0.305	Black	0.98 (0.82,1.18)	0.853
	Other	0.73 (0.60,0.88)	0.001	Other	1.26 (0.96,1.67)	0.098

a. Interaction term with marital status included in overall hazard model

b. Interaction term with marital status included in conditional hazard model

Appendix

Last Name: Barkley
SEER ID: 12082-Nov2012
Request Type: Internet Access

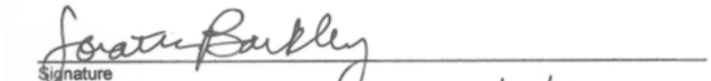
SURVEILLANCE, EPIDEMIOLOGY, AND END RESULTS PROGRAM
Data-Use Agreement for the SEER 1973-2010 Research Data File

It is of utmost importance to protect the identities of cancer patients. Every effort has been made to exclude identifying information on individual patients from the computer files. Certain demographic information - such as sex, race, etc. - has been included for research purposes. All research results must be presented or published in a manner that ensures that no individual can be identified. In addition, there must be no attempt either to identify individuals from any computer file or to link with a computer file containing patient identifiers.

In order for the Surveillance, Epidemiology, and End Results Program to provide access to its Research Data File to you, it is necessary that you agree to the following provisions.

1. I will not use - or permit others to use - the data in any way other than for statistical reporting and analysis for research purposes. I must notify the SEER Program if I discover that there has been any other use of the data.
2. I will not present or publish data in which an individual patient can be identified. I will not publish any information on an individual patient, including any information generated on an individual case by the case listing session of SEER*Stat. In addition, I will avoid publication of statistics for very small groups.
3. I will not attempt either to link - or permit others to link - the data with individually identified records in another database.
4. I will not attempt to learn the identity of any patient whose cancer data is contained in the supplied file(s).
5. If I inadvertently discover the identity of any patient, then (a) I will make no use of this knowledge, (b) I will notify the SEER Program of the incident, and (c) I will inform no one else of the discovered identity.
6. I will not either release - or permit others to release - the data - in full or in part - to any person except with the written approval of the SEER Program. In particular, all members of a research team who have access to the data must sign this data-use agreement.
7. I will use appropriate safeguards to prevent use or disclosure of the information other than as provided for by this data-use agreement. If accessing the data from a centralized location on a time sharing computer system or LAN with SEER*Stat or another statistical package, I will not share my logon name or password with any other individuals. I will also not allow any other individuals to use my computer account after I have logged on with my logon name and password.
8. For all software provided by the SEER Program, I will not copy it, distribute it, reverse engineer it, profit from its sale or use, or incorporate it in any other software system.
9. I will cite the source of information in all publications. The appropriate citation is associated with the data file used. (Please see either Suggested Citations on the SEER*Stat Help menu or the Readme.txt associated with the ASCII text version of the SEER data.)

My signature indicates that I agree to comply with the above stated provisions.



Signature

08/26/13 08/26/13

DATE Date

Please print, sign, and date the agreement. Send the form to The SEER Program:

- **By fax to 301-680-9571**
- **Or, e-mail a scanned form to seerfax@imsweb.com**

Last Name: Barkley | SEER ID: 12082-Nov2012 | Request Type: Internet Access