Ethnic Disparities in Conditional Survival of Patients with Non-small Cell Lung Cancer

Samuel J. Wang, MD, PhD,* C. David Fuller, MD,† and Charles R. Thomas Jr., MD*

Purpose: Conditional survival (CS) is an accurate estimate of survival probability for patients who have already survived at least 1 year after diagnosis. The purpose of this study was to determine whether ethnicity plays a role in 5-year CS rates for patients with non-small cell lung cancer (NSCLC).

Materials and Methods: Using the Surveillance, Epidemiology, and End Results database, we analyzed 96,480 patients with NSCLC diagnosed between 1988 and 1995. Patients were divided into five ethnic groups: White (non-Hispanic), Hispanic, African American, Asian/Pacific Islander, and Native American/Alaskan. Using the life table method, we computed observed 5-year CS rates for patients who had already survived up to 5 years after diagnosis. Results were analyzed by stage, sex, age, and histology.

Results: In general, 5-year CS rates increase for all ethnicities as time from diagnosis increases, but African Americans continued to have lower CS rates compared with other ethnic groups, even up to 5 years from diagnosis. When analyzed by stage, Hispanics with stage IV disease showed the greatest improvement in CS rate, increasing to 73% at 5 years from diagnosis. Among patients older than 70 years, African Americans had the lowest CS at 5 years—only 28%, compared with 40% to 47% for other groups. When analyzed by histology, Hispanics with large cell carcinoma had the worst CS rate (35%) at 5 years.

Conclusion: For patients with NSCLC surviving a period of time after diagnosis, 5-year CS rates vary by ethnicity. CS can provide accurate prognostic information for patients with NSCLC who have already survived several years after diagnosis.

Key Words: Survival, Ethnicity, Race, Lung cancer, Non-small cell lung cancer.

Survival estimates for patients with cancer are most often reported in the literature only as survival from the time of diagnosis. Survival probability changes, however, for patients who survive a period of time after diagnosis. Estimates of survival probability made at the time of diagnosis are less meaningful for patients who have already survived at least 1 year. Conditional survival (CS) is an accurate estimate of survival probability for patients who have survived at least 1 year after diagnosis. It is taken from the concept of conditional probability and accounts for the fact that hazard rates can change over time.

The concept of CS has important practical clinical value to both clinicians and patients. This information would allow the clinician to more accurately answer the question of a patient’s current prognosis, given that he or she has already survived a period of time since diagnosis and treatment. If the CS of a patient with cancer has increased enough to essentially match the expected survival of the general population, it allows a more objective basis to deem a patient “cured” of the disease. By relaying information about changing risk over time, CS could also be used in helping to determine the appropriate frequency of follow-up visits and aggressiveness of surveillance testing based on the patient’s current risk profile.

Several authors have previously published studies on CS for various disease sites, including breast, colon, central nervous system, lung, and other advanced carcinomas. We have also previously presented our work on CS for various disease sites at several national meetings.

In recent years, there has been greater awareness about ethnic disparities in cancer diagnosis, treatment, and prognosis. In 2003, the Agency for Healthcare Research and Quality (AHRQ) published its first annual National Healthcare Disparities Report, which highlighted the differences in access and use of health care services among various ethnicities and socioeconomic groups. The purpose of this study was to determine the effect of ethnicity in the CS of patients with non-small cell lung cancer (NSCLC).

METHODS

Definition of Conditional Survival

CS is derived from the concept of conditional probability in biostatistics. CS can be calculated from traditional actuarial or life table survival data. The mathematical definition of CS can be expressed as follows:

Let $S(t)$ be traditional life table survival at time $t$. Conditional survival, $CS(y | x)$, is the probability of surviving...
Given that the person has already survived $x$ years, conditional survival can be expressed as:

$$CS(y|x) = \frac{S(x+y)}{S(x)}$$

For example, to compute the 5-year CS for a patient who has already survived 2 years, take the life table survival at 5 years, $S(7)$ and divide this by the life table survival at 2 years, $S(2)$. As patients survive longer periods of time after diagnosis, changes in their CS are a reflection of the variation in hazard rate over time. This can be depicted graphically by examining how the slope of the traditional life table curve changes over time (Figure 1).

**Data Analysis**

The Surveillance, Epidemiology, and End Results (SEER) Program\(^2\) from the National Cancer Institute is a population-based cancer registry covering approximately 26\% of the United States population across several disparate geographic regions. When compared with the total United States population.
FIGURE 2. Ten-year overall survival for all patients with non-small cell lung cancer, grouped by ethnicity.

FIGURE 3. Five-year conditional survival for all patients with non-small cell lung cancer, grouped by ethnicity.
States population of various ethnic groups as per the 2000 United States Census, SEER coverage includes 23% of all African Americans in the United States, 40% of Hispanics, 42% of Native Americans and Alaska Natives, 53% of Asians, and 70% of Hawaiian/Pacific Islanders. SEER program registries collect data on patient demographics, cancer type and site, stage, first course of treatment, and follow-up vital status. The SEER Program obtains fairly accurate and complete mortality data by collecting information on all deaths occurring in the United States from the National Center for Health Statistics on an annual basis.

Using the April 2005 release of the SEER 13 database, we analyzed survival data from all patients diagnosed with NSCLC between 1988 and 1995. Patients were limited to those diagnosed after 1988 because this was the earliest year that SEER cases were staged with American Joint Committee on Cancer (AJCC) tumor, node, metastasis staging. Cases diagnosed up to 1995 were included to ensure at least 7 years of follow-up data (through December 2002). Cases were not included in the analysis if they were only identified by death certificate or autopsy, and only cases in which NSCLC was the first primary tumor were included.

To examine the effect of ethnicity on CS, all patients were categorized into the following five ethnic groups, based on the SEER Race Recode B variable: Caucasian (non-Hispanic), Hispanic, African American, Asian/Pacific Islander, and Native American.

![Figure 4](image-url) Five-year conditional survival for patients with non-small cell lung cancer, by stage and ethnicity. *Results suppressed for small groups in which the standard error was >10%.
lander (including Japanese, Chinese, Filipino, and Hawaiian), and Native American/Alaskan Native.

Patients who had squamous cell carcinoma, adenocarcinoma, large cell carcinoma, or any other unknown or undifferentiated carcinoma (excluding small cell carcinomas, sarcomas, or lymphomas) were included. Classification of NSCLC histologies by ICD-O-3 code was as follows: squamous cell (8052, 8070–8079), adenocarcinoma (8140–8149, 8250–8269, 8310, 8323, 8480–8481, 8550–8579), large cell (8012–8015, 8020–8029), and other/unknown (8000–8011, 8016–8019, 8030–8051, 8053–8069, 8080–8139, 8150–8249, 8270–8309, 8311–8322, 8324–8479, 8482–8549).

Using the SEER*Stat26 software with the actuarial life table method, we computed relative 5-year CS as the probability of surviving an additional 5 years after a patient had already survived from 0 to 5 years after diagnosis. We computed relative 5-year CS by ethnicity and analyzed results by AJCC stage, age >70 years, sex, and histology. All CS rates are reported as relative survival, defined as the observed survival rate divided by the expected survival rate for similar individuals matched on age, sex, ethnicity, and date at which the age was coded. Expected survival rates were taken from the SEER United States Expanded Races (1970, 1980) expected rate table. Relative survival analyses are often per-
formed to account for competing causes in SEER analyses because cause of death information is often incomplete or unknown.

RESULTS

A total of 96,480 patients with NSCLC met the inclusion criteria and were included in this analysis. Patient characteristics by ethnicity are shown in Table 1. When grouped by race, 76,957 (80%) patients were Caucasian (White, non-Hispanic); 3,398 (4%) were Hispanic; 11,152 (12%) were African American; 4,605 (5%) were Asian or Pacific Islander; and 368 (0.4%) were Native American or Alaskan Natives. Patients were further divided by stage, sex, age, and histology.

The traditional actuarial life table 10-year overall survival plot, grouped by ethnicity, is shown in Figure 2. These 10-year survival data were used to calculate the 5-year CS rates shown in Figures 3 to 7. Error bars depict the standard error for each CS statistic. Results were suppressed if the standard error was greater than 10%.

African American patients had lower CS rates compared with other ethnic groups, even up to 5 years from diagnosis (Figure 3). When analyzed by stage (Figure 4), Native Americans/Alaskan Natives with stage I disease ini-
tially had lower CS rates compared with the other groups, although this difference diminished as time progressed since diagnosis. Hispanics with stage IV disease showed the greatest improvement in CS rate, increasing to 73% at 5 years since diagnosis, compared with 39% to 56% for other ethnic groups. When examined by sex (Figure 5), there was no noticeable difference among ethnic groups. Among patients older than 70 years of age (Figure 6), African Americans had the lowest CS rates, initially increasing to 33% by year 4 but then decreasing to 28% by year 5. When analyzed histologically (Figure 7), African Americans with squamous cell carcinoma had lower CS rates than Whites or Hispanics, but this difference was not as noticeable for patients with adenocarcinoma. Hispanics with large cell carcinoma had the worst CS rate (35% at 5 years).

**DISCUSSION**

CS offers several advantages over more commonly reported static survival statistics, such as 5-year overall survival. By describing changing risk over time, it conceptually reiterates the truism that patient risk is not static. This information is potentially of great interest to both patients and clinicians. For patients who are being seen in the follow-up
Several years after their initial diagnosis, these CS tables can be used to more accurately portray their current prognosis. Psychologically, the ability of patients to more accurately conceptualize their changing risk profile may be of great benefit. Clinicians can also use CS data to implement more evidence-driven approaches to post-therapy surveillance. For instance, many physicians taper scheduled follow-up after 3 to 5 years, often with little justification based on survival data. More complete assessment of the changing profile of patient risk over time using CS provides an avenue to more cogently determine optimal intervals for follow-up visits and surveillance testing.

Additionally, CS can be used to detect differentials in population survival patterns. For instance, crude 5-year DSS does not account for the fact that the vast majority of mortality risk may fall earlier or later in the span of time assessed. Like Kaplan-Meier graphs, CS rates provide a ready mechanism to assess differential temporal risk of mortality in a given population.

Using the SEER dataset represents an effort to determine CS parameters based on large population cohorts. It should be noted, however, that increased cohort size comes at the cost of treatment homogeneity. The SEER dataset, while geographically limited, represents the best large-scale pool of...
patient data, and allows us to make reasonable estimates of CS rates that are broadly applicable for the general United States population.

CS in lung cancer has been previously explored in other series, although the application of ethnicity as a covariate has not been previously reported. Compared with these prior studies, our overall CS results are consistent with those of both Merrill et al., who used an earlier version of the SEER database, and Skuladottir et al., who analyzed patients from the Danish Cancer Registry. Perhaps not surprisingly, Merrill et al.’s results most closely correspond to our results for White patients because Whites account for 80% of patients in SEER.

That ethnic populations exhibit survival differentials has been established in numerous series. Similar to our study, several of these investigators have also found that African Americans have worse survival at diagnosis compared with other ethnicities. Our study also shows that, whereas CS generally improves over time for most groups of patients, African Americans often continue to have worse CS rates compared with Whites up to 5 years after diagnosis, particularly in earlier stage disease, male patients, age >70
years, and squamous cell histology. For patients with advanced stage disease or large cell histology or who were female, however, the CS rate for African Americans shows greater improvement over time and actually exceeds that of Whites as time elapses from diagnosis.

In most subsets, we found that the Native American/Alaskan Native group had the poorest initial survival but generally showed a large improvement in CS over time. Hispanics seemed to do relatively well in stage IV disease but relatively poorly if they had large cell histology. However, the numerical paucity of patients in these ethnic subgroups and the incumbent statistical uncertainty thereof precludes ready comparison with the much larger ethnic categories reported.

To determine whether the apparent ethnic differences in CS could be explained solely by differences in other confounding factors, we performed a multivariable regression analysis to simultaneously control for other covariates. We constructed a series of Cox proportional hazard models (JMP 6 Software, Cary, NC) on subsets of the survival data 1 to 5 years after initial diagnosis, using covariates of stage, age >65 years, sex, histology, and race to determine the independent contribution of race to differentials in CS. In this multivariate analysis, we found that the effect likelihood ratio for the race category was statistically significant for every year after diagnosis, except for year 5. Although the actual risk ratio between any two races varied in statistical significance from year to year, this analysis showed that race is still an independent prognostic factor for CS at least 4 years after diagnosis. Our findings are consistent with what others have reported.

We have found that CS in lung cancer varies by ethnicity, and these data can provide more accurate prognostic information for patients who have already survived several years since diagnosis.

As in other studies of CS, patients with advanced stage disease and poorer initial prognoses have the greatest increases in CS as they survive for longer periods of time from diagnosis. The rapid increase in CS for advanced disease may have useful clinical application. For patients with stage IV disease who survive >3 years from diagnosis, subsequent mortality outcomes are comparable to those of patients with stage I disease at diagnosis. Consequently, if one follows-up patients with stage I disease for x years, patients with stage IV disease should be similarly followed for x + 3 years. Using this type of CS information can allow optimal follow-up testing frequency and duration to be based on a patient’s actual disease risk, rather than simply on custom or tradition.

The SEER Race Recode B variable was used to determine ethnicity. SEER uses medical records and registration information to determine all races, except for Hispanic, which is determined by surname. Whereas the designation of Whites and African Americans in SEER has been studied and found to be reliable, the designation for other ethnicities has not been externally validated, so its accuracy is unknown. It has been hypothesized that the use of surnames to determine Hispanic race may result in misclassification of more women than men because of marital name changes; nevertheless, preliminary studies indicate that the race variables in SEER are still likely more accurate than those in Medicare databases.

Throughout this analysis, relative survival rates, defined as observed survival divided by the expected survival for that patient, are reported. Because cause of death information is often unreliable in the SEER database, relative survival is a useful alternative measure that assesses the proportion of excess deaths that occur among patients with lung cancer compared with the general population. It serves to adjust for differences in observed survival that may be the result of competing causes of death. Expected survival rates were obtained from the SEER 1970, 1980 Expanded Races expected rate table by matching patients by age, sex, race, and the date at which age was coded.

It is imperative for physicians to more specifically characterize, use, and communicate patient risk profiles as accurately as possible. Models such as Adjuvant! Online or MSKCC risk nomograms are especially inviting as a mechanism for initial risk stratification. Incorporation of a CS component would allow patients and physicians to re-evaluate the patient’s risk as it changes over time, which would potentially afford more accurate management of disease as a part of the physician’s surveillance algorithm. Similarly, ethnicity, which is a key component of risk stratification at diagnosis, should be implemented as a covariate in CS-based analyses.

Finally, efforts must be made to successfully communicate this changing risk profile to the patient, an often difficult task for the oncologist. When patients that are seen in follow-up inquire about their current prognosis, they should be given an accurate risk assessment that accounts for the time already survived since diagnosis.

We have found that CS in lung cancer varies by ethnicity, and these data can provide more accurate prognostic information for patients who have already survived several years since diagnosis.

REFERENCES

8. Skuladottir H, Olsen JH. Conditional survival of patients with the four
major histologic subgroups of lung cancer in Denmark. J Clin Oncol

9. Kato I, Severson RK, Schwartz AG. Conditional median survival of
patients with advanced carcinoma: surveillance, epidemiology, and end

10. Wang SJ, Luh JY, Thomas CR. Conditional survival of breast cancer
patients. American Society of Breast Diseases, 29th annual symposium.
Las Vegas, NV, April 14, 2005.

patients. American Radium Society, 87th annual meeting. Barcelona,
Spain, April 30, 2005.

12. Wang SJ, Luh JY, Thomas CR. Conditional survival of patients with

13. Fuller CD, Wang SJ, Thomas CR. Conditional survival of gallbladder
adenocarcinoma treated with radiotherapy: Analysis from the SEER
database. American Society for Therapeutic Radiology and Oncology,

14. Wang SJ, Luh JY, Fuller CD, Thomas CR. Impact of ethnicity on
conditional survival of breast cancer patients: Analysis from the SEER
database. 28th Annual San Antonio Breast Cancer Symposium, San
Antonio, TX, December 8, 2005.

15. Wang SJ, Luh JY, Thomas CR. Conditional survival of ovarian cancer
patients. Aung Young Investigators Research Forum, Houston, TX,
September 12, 2005.

patients. American Society of Clinical Oncology Prostate Cancer Sym-

17. Wang SJ, Fuller CD, Luh JY. Conditional survival of patients with
gastric adenocarcinoma: A SEER database analysis. Society of Surgical

An analysis from the SEER database of conditional survival for gastric
adenocarcinoma. American Society of Clinical Oncology Gastrointesti-

patients with colon cancer: An analysis of National Surgical Adjuvant
Breast and Bowel Project (NSABP) trials C-03 through C-06. American
Society of Clinical Oncology, 42nd Annual Meeting Proceedings, At-
tlanta, GA, June 2–6, 2006.

20. Luh JY, Wang SJ, Fuller CD, Thomas CR. A SEER database analysis of
conditional survival for prostate cancer patients. American Society of
Clinical Oncology, 42nd Annual Meeting Proceedings, Atlanta, GA,
June 2–6, 2006.

carcinoma: Results from the SEER 11 dataset. American Society of
Clinical Oncology, 42nd Annual Meeting, Atlanta, GA, June 2–6, 2006.

ability for head and neck squamous cell cancer: Results from the SEER
11 dataset. American Society of Clinical Oncology, 42nd Annual Meet-
ing Proceedings, Atlanta, GA, June 2–6, 2006.


24. Surveillance, Epidemiology, and End results (SEER) Program;
Sub (1973–2002 varying), National Cancer Institute, DCCPS, Surveil-
lance Research Program, Cancer Statistics Branch, released April 2006,
based on the November 2004 submission. Available at: http://www.seer.

25. American Joint Committee on Cancer. AJCC cancer staging manual,

26. Surveillance Research Program, National Cancer Institute SEER*Stat

27. Lee ES, Roberts RE, Labarthe DR. Excess and deficit lung cancer


29. Cooley ME, Jennings-Dozier K. Lung cancer in African Americans: a

Impact of race in lung cancer: analysis of temporal trends from a
surveillance, epidemiology, and end results database. Chest 2001;120:
55–63.

African-American/non-African-American patients with advanced non-
small-cell lung carcinoma: report from the Cancer and Leukemia Group

32. Clegg LX, Li FP, Hankey BF, Chu K, Edwards BK. Cancer survival
among US whites and minorities: a SEER (Surveillance, Epidemiology,
and End Results) Program population-based study. Arch Intern Med


Limited access to care: effect on racial disparity and prognostic
factors in lung cancer. Cancer Epidemiol Biomarkers Prev 2006;15:

36. Flenaugh EL, Henrichs-Forsythe MN. Lung cancer disparities in Afri-
439.

37. Willisie SK, Foreman MG. Disparities in lung cancer: focus on Asian
Americans and Pacific Islanders, American Indians and Alaska Natives,

38. Thun MJ, Henley SJ, Burns D, Jemal A, Shank T, Calle EE. Lung
cancer death rates in lifelong nonsmokers. J Natl Cancer Inst 2006;98:
691–699.

demographic and socioeconomic characteristics in the SEER-Medicare


42. Memorial Sloan-Kettering Cancer Center. Prediction tools 2006. Avail-

43. Kattan MW. Nomograms are superior to staging and risk grouping
systems for identifying high-risk patients: preoperative application in

44. Van Zee KJ, Manasseh DM, Bevilacqua JL, et al A nomogram for
predicting the likelihood of additional nodal metastases in breast cancer
patients with a positive sentinel node biopsy. Ann Surg Oncol 2003;10:
1140–1151.

45. Diblasio CJ, Kattan MW. Use of nomograms to predict the risk of
disease recurrence after definitive local therapy for prostate cancer.

for patients undergoing resection for adenocarcinoma of the pancreas.