2016

DEVELOPMENT OF MULTIPLE PRIMARY CANCERS IN LUNG CANCER PATIENTS: APPALACHIAN VS. NON-APPALACHIAN POPULATIONS OF KENTUCKY

Vira Pravosud
University of Kentucky

STUDENT AGREEMENT:

I represent that my capstone and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's capstone including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Vira Pravosud, Student

Thomas Tucker, PhD, Committee Chair

Corrine Williams, ScD, MS, Director of Graduate Studies

DEVELOPMENT OF MULTIPLE PRIMARY CANCERS IN LUNG CANCER PATIENTS: APPALACHIAN VS. NON-APPALACHIAN POPULATIONS OF KENTUCKY

CAPSTONE PROJECT PAPER

A paper submitted in partial fulfillment of the requirements for the degree of
Master of Public Health

in the

University of Kentucky College of Public Health

By

VIRA PRAVOSUD

Lexington, Kentucky
April 20, 2016

Committee Members

_____
Thomas Tucker, PhD Chair

_____
Glyn Caldwell, MD, PhD

_____
Bin Huang, DrPH

_____
Nathan Vanderford, PhD

CONTENTS

ABSTRACT

OBJECTIVE: This study examined whether there were differences in the development of multiple primary cancers in lung cancer patients residing in the Appalachian versus Non-Appalachian regions of Kentucky. The study also identified other factors associated with the development of another primary cancer in lung cancer patients.

HYPOTHESIS: Lung cancer patients residing in Appalachian Kentucky are more likely to develop multiple primary cancers compared to patients residing in Non-Appalachian Kentucky.

METHODS: This was a retrospective, population-based cohort study of Kentucky patients (N=26456) aged 20 years and older, diagnosed with primary lung cancer between January 1, 2000 and December 31, 2013. The study population was drawn from the Kentucky Cancer Registry. Cases were excluded if they were diagnosed with second primary cancers within 3 months after the diagnosis of their first primary lung cancer, and if they changed their residence moving from Appalachian to Non-Appalachian region or vice versa. Subjects were followed to determine if they developed subsequent primary cancers. The Cox proportional hazards model was used to control for the time from diagnosis to death or a second PC.

RESULTS: The final adjusted multivariable hazards model indicated that there were no statistically significant differences between Kentucky Appalachian and Non-Appalachian lung cancer populations with respect to the hazards of developing of a subsequent primary cancer (HR: 1.002, p=0.9713). The adjusted analysis revealed that increasing age at diagnosis, male gender, and patients having surgery increased the hazards of developing another primary cancer (HR: 1.015, p=0.0001; 1.169, p=0.012; 1.446, p=0.0003). Having a stage IV tumor decreased the hazards of the outcome by 31.6% comparing to the patients with stage II tumors (HR=0.684, p=0.0015).

CONCLUSION: No differences were found between Appalachian and Non-Appalachian lung cancer patients. Surgery was very likely associated with getting a second primary because patients who had surgery were likely to live longer, and thus, had a greater opportunity to develop a second primary. In contrast, patients who were diagnosed with stage IV lung cancer had very short survival times and were, thus, less likely to develop a second primary cancer.

I. INTRODUCTION

1.1. Objective of the Study.

According to America's Health Ranking, Kentucky is ranked 44th compared to all other U.S. states [1]. Kentucky residents have a high prevalence of smoking, obesity, diabetes, heart diseases, and cancers. The incidence and mortality rates for lung, breast, cervical and colorectal cancers are particularly high [2]. In 2012, Kentucky had the highest age-adjusted cancer incidence and mortality rates in the country compared to other U.S. states (515.1 and 201.2, per 100,000 population respectively) [3]. During the time period 2000 through 2013 the age-adjusted incidence and mortality rates for all cancers in the state of Kentucky were (561.5 and 212.2). Significant differences in age-adjusted cancer incidence rates for the 2000-2013 time period can be observed between Kentuckians residing in Appalachian and Non-Appalachian regions (529.8 versus 521.5) [4]. Cancer mortality rates also differ significantly between the Appalachian and Non-Appalachian areas of the state. Namely, age-adjusted cancer mortality rates for the period from 2000 through 2012 were much higher for those who reside in Appalachia Kentucky comparing to the Non-Appalachian area of the state (225.8 versus 191.6) [5].

The existing literature postulates that improved clinical surveillance, including early screening and detection, has contributed to a decrease in cancer mortality. However, there has also been a notable increase in multiple primary malignancies among cancer patients, a phenomenon that has captured a lot of attention from clinicians and epidemiologists [6-8]. A number of factors may influence the risk of developing subsequent primary cancers, including genetic predisposition, environmental exposures, and the life style of individuals [6-13].

For this study, Kentucky lung cancer patients were defined as the target population. Lung cancer is the leading cause of death and the second most commonly occurring cancer in both

men and women in the U.S. [14, 15]. Lung cancer incidence and mortality rates are notably higher in Kentucky than nationwide [16]. During 2007-2011, the age-adjusted incidence rate for lung cancer in Kentucky was 122.9 for men and 80.7 for women. These rates are higher than the estimated incidence rates for the nation as a whole during the same time period for men and women, respectively (78.6 and 54.6) [16]. The age-adjusted lung cancer mortality rates during 2007-2011 were higher in Kentucky than in the US for both men and women, respectively (94.5 and 55.5 vs. 61.6 and 38.5) [16].

Within the state of Kentucky rates differ with respect to the region of residence. Specifically, during 2000-2013 the age-adjusted lung cancer incidence and mortality rates in the Appalachian region were higher than the rates in the Non-Appalachian region. The age-adjusted incidence rate in the Appalachian region was 109.8 per 100,000 persons, whereas the estimated incidence rate for lung cancer in the Non-Appalachian region for the same time period was 95.3 per 100.000 persons [17]. The age-adjusted lung cancer mortality rate in the Appalachian region during 2000-2012 was 84.8 per 100,000 persons, whereas the estimated mortality rate for lung cancer in the Non-Appalachian region during this same time period was 64.9 per 100.000 persons [18].

Smoking and exposure to environmental carcinogens are associated with an increased risk in lung cancer. Additionally, interactions of these exposure factors with a genetic predisposition may significantly increase the odds of developing a first lung cancer, and then another subsequent primary cancer [6, 19, 20]. Therefore, considering the high incidence and mortality rates of cancers in the state, and the existing environmental, socio-economic, and health-related differences between the Appalachian and Non-Appalachian regions, this study aimed to determine whether lung cancer patients from the Appalachian Kentucky developed subsequent

(second) primary cancers more frequently than lung cancer patients residing in the Non-Appalachian area of the state. Also, this study tried to examine other factors associated with the development of subsequent primary cancers in patients first diagnosed with lung cancer.

1.2. Background.

This study did not focus on examining subsequent primary cancers of the lung, but rather on development of subsequent primary cancers of any type, including secondary cancers of lung. Determining whether multiple lung tumors occur independently or due to the metastases has been an issue in lung cancer pathology. Studies describe cases of misdiagnosis of independent primaries appearing at the same time and tumors occurring due to metastases [21]. There are several reasons why the diagnosis of a second primary lung cancer is considered to be a complex issue. Difficulties exist in differentiating between metastatic and primary lesions if they occur within 2 years of the initial tumor. Differential diagnostics is also difficult if both the second primary lesion and initial tumor are of a similar histological subtype, and/or if the second lesion is located in the area of previous radiotherapy, since the latter leads to the changes in the tissue morphology [22].

The results of a voluntary survey of the Pulmonary Pathology Society showed that pathologists use different approaches to identify whether two anatomically distinct foci of lung carcinoma are intrapulmonary metastases or independent primaries [23]. Several studies described different methods of distinguishing whether multiple lung cancers occur due to metastases or patients have independent (in most cases, synchronous) primaries [21, 24, 25]. For instance, the diagnostic lineage test based on genomic rearrangements from mate-pair sequencing [24], the histologic-mutational methods, and the driver-mutational testing of selected

genes are recommended to be used for diagnosis of non-small synchronous cell lung cancers in patients with lymphatic metastases [26].

However, in most cases, a traditional method of diagnosis, known as Martini and Melamed criteria, is used. It is based on identifying tumor characteristics, i.e. morphology, location, presence or absence of carcinoma in situ, vascular invasion and metastasis. A number of studies described the increasing incidence of multiple primary lung cancers and issues related to their diagnosis, treatment and survival [26-38]. Previous research showed that the risk of developing multiple primary lung cancers ranges between 1 and 15 percent [22]. In most cases, multiple lung tumors are presented as secondary primary tumors, which can be synchronous or metachronous. It was indicated that synchronous multiple lung cancers had high occurrence in patients with idiopathic pulmonary fibrosis, who had the following characteristics: (1) male patients, (2) smokers, (3) had small cell carcinomas of (4) peripheral type in (5) lower lobes [27]. In contrast, according to Bhaskarla et al., a small proportion of lung cancer patients develop second primary lung cancers either synchronous or metachronous. In those patients, the reoccurrence was related to female gender, younger age, earlier stage, and white race [8].

The risk of developing multiple primary cancers has been explored in various studies. Cancer patients can have multiple primary cancers detected in different sites, or they can occur in the same organ due to various exposures or genetic predisposition [39]. One study showed that there was a 2.2 times greater risk of developing a concordant secondary neoplasm, and 1.1 risk of developing a metachronous tumor of a different type and in another site [9]. The best explanation for these differences is that development of subsequent primary malignancies depends on the site of cancer as well as interactions of genetic predisposition and exposures that people may have [6].

The results of the analysis based data from the Surveillance, Epidemiology, and End Results (SEER) database have shown that the occurrence of multiple primary cancers varies depending on the site of origin. Namely, it ranges from the lowest 1 percent for an initial liver primary diagnosis to the highest of 16 percent for initial bladder cancer primaries. The highest percentages of reoccurrence or development of subsequent cancers of other sites were observed in patients with first primary cancers of urinary bladder, oral cavity and pharynx, kidney and pelvis, colon and rectum, melanoma of skin, prostate, and breast [7].

According to these results, four percent of the patients who had the first primary cancer of the lung or other respiratory organs (including bronchus, trachea, mediastinum, and other organs) developed subsequent primary cancers. The median age at diagnosis of lung or other respiratory organs' cancers was 67 years, and the 5-year observed survival rate for those lung cancer patients who developed multiple primaries was 11 percent [7].

Previous research showed that a number of factors are related to development of multiple primary cancers. They include effects from tobacco and alcohol consumption, infections, immunosuppression, genetic predisposition, and also effects of treatment of the initial cancer, the role of drug-metabolizing enzymes, DNA repair proteins, and drug pharmacokinetics [6, 10].

For instance, a systematic review of studies and meta-analysis looking at occurrence of second primary malignancy after radioactive iodine treatment for thyroid cancer revealed an increased risk of subsequent cancers (RR=1.19, 95% CI: 1.04-1.36) in those patients who received radioactive iodine treatment [40]. In general, the increased risk of such cancers as osteosarcomas, melanomas, and soft-tissue sarcomas is related to previously received radiotherapy [6].

The association of surgical procedure and hazards of occurrence of another primary cancer has not been assessed by previous research. At the same time, it was found that aggressive surgical approach was beneficial for lung cancer patients with multiple synchronous lung lesions, and could positively affect patients' survival rate [25, 28]. As indicated in the literature, development of subsequent primary cancers was positively associated with survival time (with the exception of thyroid cancer). Overall, the investigators concluded that, in order to develop multiple primary cancers, patients had to demonstrate relatively long survival after their initial diagnosis [7].

Several factors were found to be related to the observed survival of cancer patients. Research shows that in lung cancer patients with non-small cell carcinomas, the risk of reoccurrence of cancer in the same site or development of a subsequent primary in another organ is not associated with a specific histologic type of tumors (i.e. squamous, adenocarcinomas) [41]. However, it was demonstrated that lung cancer patients with adenocarcinomas had better survival than patients with tumors of other histologic type, especially those with stage II cancers [42]. Also, female lung cancer patients with stage II tumors are known to have better survival compared to men [7]. This can be explained by a higher frequency of surgical procedures performed in women lung cancer patients, especially those with stage II tumors [7, 42].

In fact, radiotherapy also increases survival in locally advanced non-small cell lung cancer patients who cannot tolerate surgical or drug treatments, especially if radiotherapy was combined with chemotherapy [43].

Associations of smoking and occurrence of second primary cancers have been discussed in the existing literature [6, 7]. A positive association of smoking and second primary cancers was identified among patients with head and neck squamous cell cancers [19]. Assessment of

standardized incidence ratios (SIRs) of subsequent primary cancers in all pairs of smoking-associated cancers revealed that SIRs were much higher for women. This study also indicated that, in general, the magnitudes of the ratios are greater due to strong negative effects of smoking [20].

Race and gender were found to be significant predictors of occurrence of subsequent primary cancers, with females and blacks having demonstrated larger odds of developing another primary cancer [6]. Namely, blacks had a higher risk of developing a second lymphoproliferative malignancy in patients with multiple myeloma [11], as well as ipsilateral or contralateral breast cancers [13]. However, whites were more likely to develop second primary lung [8], endometrial, and ovarian cancers [13], and both blacks and whites had greater odds of developing metachronous tumors as compared to representatives of other races among renal cell carcinomas survivors [12].

Age is also an important factor. Koubkova et al., in a literature review, refer to the studies concluding that a huge burden of multiple primary malignancies was observed in the group of patients aged 30-59 years [6, 44], whereas one Czech study demonstrated that the highest burden of occurrence of subsequent cancers was observed in the older group – 50-69 years [6, 10]. Interestingly, patients who were cancer survivors during their childhood, had 6 times greater relative risks of developing a second primary cancer [6, 44].

Therefore, we can note that a number of studies have explored the phenomenon of multiple primary malignancies in different cancer populations. However, there still is a need to examine various factors that may predict the occurrence of subsequent primary cancers in lung cancer patients, as this issue is not substantially described in the available literature.

1.3. Public Health Importance and Impact of the Study.

No previous study has been conducted to compare whether there are differences in development of multiple primary cancers in lung cancer patients residing in Appalachian and Non-Appalachian regions of Kentucky. The following research project is intended to address this question, and to describe other factors that impact and can predict the development of subsequent primary cancers in patients first diagnosed with lung cancer.

1.4. Research Questions and Hypotheses.

The objective of this study was to determine whether there were differences in the development of subsequent primary cancers among residents in Appalachian and Non-Appalachian Kentucky who were first diagnosed with lung cancer, and assess which biological and demographic factors, as well as types of treatment (surgery or radiotherapy) that Kentucky lung cancer patients received, were associated with the development of another primary cancer.

The primary hypothesis of the study was that lung cancer patients residing in Appalachia Kentucky were more likely to develop multiple primary cancers compared to patients residing in Non-Appalachia Kentucky. The reasoning behind this assumption was based on the fact that Appalachian Kentucky is characterized by a high prevalence of smoking rates, especially among males, as well as occupational and environmental exposure to coal-mining, low socio-economic status, lack of commercial health insurance coverage, and health professionals shortages [45, 46].

1.5. Delimitations of the Study.

This study had several delimitations.

Firstly, the SEER database includes information about Educational Attainment, poverty level, and employment which is presented in the form of calculated percentages or rates based on Census tracts [47]. Since there was no individual level data, the investigators decided not to

include these variables in the analysis. Further research is recommended to assess the effect of these variables as potential residual confounders.

Secondly, due to the nature and source of data, there were many missing values. The investigators decided to set those missing cases to separate groups, which were also included in the analysis. This was done in order not to exclude a large number of subjects, which would mean a loss of power. Imputing was not considered as an appropriate approach to address the issue of missing values for this particular study.

Thirdly, increased incidence rates of many cancers are known to be attributed to older age [48]. As a result, a number of researchers either use stratification by age, or treat age as a categorical variable, so the differences between different age groups can be observed even assuming the loss of power for every additional cell. The range of age at which patients were diagnosed with their initial lung cancer was quite large, including patients from 21 to 100 years old. However, in this study, the investigators decided to treat age as a continuous variable, due to the fact that median age at diagnosis of lung cancer was 67 years, with first and third quartiles being 59 and 74 years, respectively. Thus, the majority of subjects would be included into the same age category.

Fourthly, SEER data included a variable that expressed the sequence of received treatment for the initial lung cancer, showing whether radiation or surgery went first if a patient had both. Also, SEER data included a variable which showed the type of radiation a patient received. However, the investigators decided not to assess the dose-response effect of different radiation types, and not to account for the number or order of treatments that a patient could receive. At this point, there was no support from the reviewed literature that would have allowed for an adequate hypotheses to be made. Furthermore, there were many cases with missing

treatment information. This issue could be solved by linking SEER data to Medicaid or Medicare data. Thus, further investigations are suggested in order to address whether there are any associations between development of subsequent primary cancers with the sequence of treatment and/or dose-response effects of radiation.

II. METHODOLOGY

2.1. Study design. Source of data.

This was a retrospective, population-based cohort study using the SEER database was used for this research. The SEER Program of the National Cancer Institute collects data on cancer cases from a number of registries all over the U.S. beginning in 1973. Kentucky healthcare institutions and establishments also are required to report all cancer cases to the Kentucky Cancer Registry, which has participated in the SEER Program since 2000 [49].

To explore the phenomenon of multiple primary cancers, the SEER database was considered the best option. According to the SEER rules, recurrences or tumors of the same histology which are reported within 2 months, are excluded. Thus, there was no doubt in potential misclassification of primary cancers [7].

2.2. Study subjects. Availability of data.

Kentucky patients diagnosed with lung cancer as their first or the only primary from January 1, 2000 through December 31, 2013 were considered the target population for this study. Residence in Appalachian region of Kentucky was defined as the primary exposure.

The study cohort included 47,166 of subjects with lung cancer, divided into two groups based on their residence status: Appalachian (15,780; 33.46%) or Non-Appalachian (31,386; 66.54%) region of Kentucky. Among 47,166 lung cancer patients, 44,380 (94.09%) were diagnosed with lung cancer as their only primary, whereas 2,786 (5.91%) patients had lung cancer as their first primary, but later developed another primary cancer. The outcome of interest was the development of a second primary cancer after at least 3 months from the time of the diagnosis of lung cancer. This time frame was taken into consideration with the purpose to avoid

potential misclassification of the synchronous secondary cancer and a possible misclassification with metastasis.

Identified cases were retrospectively followed during the observable survival time after their diagnosis of lung cancer in order to see if patients from Appalachian region were more likely to develop another primary cancer. In addition, this study examined other factors that could have impacted the development of subsequent cancers in lung cancer patients of Kentucky.

To find out how many patients had developed further primary cancers, another SEER dataset was requested. It contained information on 41,883 subjects who were diagnosed with second primary cancers. Based on patient identification, after merging these two datasets (total of 86,721 subjects), it became possible to find out that 2,328 subject who were initially diagnosed with lung cancer as their first primary were present in both datasets, meaning that they eventually developed another cancer, whereas 44,380 subjects had only a single primary lung cancer.

2.3. Inclusion and Exclusion Criteria.

The inclusion criteria were the following:

- Kentucky lung cancer patients, who had lung cancer as their only or the first primary cancer,
- patients aged twenty years and older,
- malignant tumors only.

The exclusion criteria were the following:

- small or large cell carcinomas,
- development of a subsequent primary cancer within 3 months,

- changing the place of residence moving from Appalachian to Non-Appalachian region or vice versa,
- being initially coded as those having two or more primaries, but not having any records of developing another primary cancer.

Therefore, this study included patients age twenty and older. All the subjects had malignant tumors. Patients (n=880) who were diagnosed with the second primary cancer within 3 months after the diagnosis of their first primary cancer were excluded from the study. Those who were diagnosed with either small- or large-cell carcinomas were excluded from the study (n=19,261).

Eleven patients who changed their residence, moving from Appalachian to Non-Appalachian region or vice versa, after their first diagnoses, were also excluded from the study population.

In addition, 631 patients who did not develop a subsequent primary cancer after being diagnosed with lung cancer as their first primary, but who were initially marked as those having two or more primaries, were excluded from the final sample. There was a high probability that those subjects were either mistakenly coded as having multiple primaries, or those patients were diagnosed with multiple primaries simultaneously to their lung cancer diagnosis, which did not satisfy the previously mentioned criteria of not developing a subsequent primary cancer within 3 months period.

2.4. Final Study Sample.

As mentioned in the section 2.2., 46,708 (2,328 who developed a subsequent primary cancer and 44,380 who did not) were initially identified from the existing datasets. However, 20,254 subjects were excluded, as described in section 2.3. Therefore, the final sample contained

26,456 subjects, 1,438 of whom had multiple primary cancers, and 25,018 had single primary lung cancers.

2.5. Data Collection and Measures. Subject Recruitment Methods.

Secondary data, including information on patient age, gender, race, histology and morphology of cancer, stage at diagnosis, and received treatment, was obtained from the SEER database. The data was limited to Kentucky cancer cases only.

This study was approved by the Institutional Review Board of the University of Kentucky (#16-0210-X1B) and received Exemption Certification for Protocol No. 16-0210-X1B.

2.6. Potential Risks for Study Subjects.

It should be mentioned here that the research involved no more than minimal risk as subjects were not recruited for this study. Patient information was earlier reported to SEER. All data used for analysis was de-identified.

2.7. Confidentiality.

Since SEER database does not include any personal identifier, i.e. patient name, social security number, personal addresses, no patients were identified. Moreover, no attempts will be made, neither in the way of potential publications nor in presentations of the results of the study, to use the data to identify those patients.

2.8. Data Analysis.

Statistical analyses were carried out using SAS version 9.4.

Exploratory analyses included the use of descriptive statistics. Specifically, means, standard deviations, quartiles, and ranges were examined for continuous variables, and frequencies and percentages were examined for categorical variables. Results were examined by

main exposure (Appalachia versus Non-Appalachia Kentucky), which was of primary interest in this study. Therefore, descriptive statistics are presented by exposure groups in Table 1.

In addition, for categorical variables, Pearson χ^2 and Fisher's Exact Tests (if the expected cell was ≤ 5) were used to determine if the associations were statistically significant. The Wilcoxon-Mann-Whitney test was used for continuous variables to determine differences between two exposed groups. Two-sided p-values $\leq .05$ were considered statistically significant.

The Wilcoxon-Mann-Whitney test was used to determine if two groups were different with regards to age, survival time, and time during which a subject could develop an event of interest. The Wilcoxon-Mann-Whitney test was carried out, since those three continuous variables failed the normality tests (Kolmohorov-Smirnov test: $P_{age}<.01$, $P_{survival}<.01$, and Shapiro-Wilk: $P_{timetoevent}<.0001$).

There were significant differences between lung cancer patients residing in Appalachia Kentucky compared to those residing in the Non-Appalachian region with respect to the number of primary cancers they developed during their lifetime as shown in Table 1. It should be emphasized here that the purpose of this research was to assess if there were differences between Appalachian and Non-Appalachian populations regarding development of subsequent primary cancers in lung cancer patients. Therefore, the aim was to assess differences among those who developed multiple primary cancers (regardless of the number of those primary cancers) compared to those who did not. Table 2 presented descriptive statistics for patients in each of the two outcome groups: those with multiple primaries against those with single primaries. Similarly to Table 1, Pearson χ^2 and Fisher's Exact Tests (if the expected cell was ≤ 5) were used to determine whether the associations for categorical variables were statistically significant. The

Wilcoxon-Mann-Whitney test was used for continuous variables to determine differences in two outcome groups. Two-sided p-values $\leq .05$ were considered statistically significant.

The next step was to assess the unadjusted associations for independent covariates and the outcome – development of multiple primary cancer. Therefore, univariate analysis using Cox proportional hazards models was carried out, with:

- development of multiple primary lung cancers as the "censor" variable;
- the "time" variable was calculated as observed survival from the time of the lung cancer diagnosis for those who did not develop another primary cancer until death or the end of the study period; and the difference in observed survival between 1st and 2nd diagnoses for those who developed another primary cancer.

The univariate and multivariable Cox proportional hazards models were used to control for the time from diagnosis to death or a second primary cancer (the differences were considered statistically significant for p-values of <0.05).

According to Table 3, which presents the results of univariate models, the main exposure covariate – residing in Appalachian region – was not statistically significantly associated with the outcome. Race, unknown status for patients' tumor stage, receiving radiotherapy treatment, surgery, and adenocarcinoma diagnosis were also not found to be significant. All other associations were statistically significant. As a result, the next step was to carry out a multivariable Cox proportional hazards regression in order to estimate adjusted hazard ratios of developing subsequent primary cancers, as well as to examine the factors associated with development of a second cancer.

It was decided to fit a multivariable Cox proportional hazards model with all of the assessed covariates with the exception of race. The association of race with the outcome was not

significant, the number of blacks with multiple primaries was small (59 cases) and there was no particular research interest to observe differences between patients of different races.

Backward elimination with a significance level of five percent was carried out for the first multivariable model, starting with the main effects and interactions for all the aforementioned potential predictors. Therefore, the first multivariable Cox proportional hazards model included the following covariates: Appalachian region, Age at diagnosis, Histologic Type ICD-O-3, Grade, Tumor Stage, Gender, Radiotherapy, and Surgical procedure. The regression revealed that there were no statistically significant associations between the outcome and Appalachian region, Histologic Type ICD-O-3, and Radiotherapy[1]. Since residence in the Appalachian region was the main exposure and receiving radiotherapy as a treatment for the initial lung cancer was of a great research interest, only Histologic Type ICD-O-3 covariate was excluded. The second model was fit with the remaining two-way interactions.

As for model fit statistics, the second model (without Histologic Type ICD-O-3 covariate) had a lower AIC comparing to the first model: AIC_1=17486.26, SBC_1=17555.837, AIC_2=17484.67, SBC_2=17544.308. Therefore, the second model was considered the final one.

In addition, to explain the study sample better, unadjusted Kaplan-Meier estimated survival curves and corresponding log-rank test results are presented in the Appendix for those covariates which were significant in the final model. Included are comparisons of (1) patients residing in Appalachian versus Non-Appalachian regions, (2) two gender groups, (3) patients diagnosed with lung cancer on a different stage and (4) with a different grade of tumor, (5) patients who had or had not had their surgery performed. The Kaplan-Meier estimated survival

[1] Appendix. Model 1

curves were used to visually show differences in survival between comparison groups, and the log-rank test was used to determine statistical significance with respect to the population survival curves.

2.9. Independent and Dependent Variables.

Variable 1 - *Appalachia Region* – patients were divided in two geographical regions: Appalachia and non-Appalachia Kentucky, depending on their official residence at the time of the diagnosis of their lung cancer [50].

Variable 2 – *Age at Diagnosis* (measured in years) – defines the age of the patient when they were diagnosed with lung cancer [50].

Variable 3 – *Gender* – patients were divided into female or male gender groups [50].

Variable 4 – *Race* – patients were divided into three groups: Blacks, Whites, and Other, which includes American Indian/AK Native, Asian/Pacific Islander, and also those with no information provided [50].

Variable 5 – *Vital status* – defines the patients' status on December 31, 2013 as the study cutoff date [50]. The variable was only used to determine survival times and in the descriptive analysis.

Variable 6 – *Number of Survival Months after Diagnosis of Lung Cancer* – defines observed survival time measured in months [50].

Variable 7 – *Time to Event (development of a subsequent primary cancer)* – defines observed time, measured in months, from data of diagnosis to the diagnosis of a second primary or death.

Variable 8 – *Number of Primaries* – defines the total number of tumors for a person developed during their lifetime [50]. The variable was used for exploratory purposes only and in the descriptive analysis.

Variable 9 – *Histologic Type ICD-O-3* **–** the records of tumor histology coded according to ICD-O-3 values [50, 51]. For this research project, only non-small (adenocarcinomas and squamous) cell carcinomas were included.

Variable 10 – *Summary Stage 2000 (1998+)* was derived from Collaborative Stage (CS) for 2004+ and extent of disease (EOD) from 1998-2003 [50]. The following were the values: in situ, localized, regional, distant, and unknown.

Variable 11 – *Grade* – defines grade of the tumor based on grade codes in ICD-O-3. (Cases diagnosed in 1973-2000 were coded in earlier versions and may lack the specificity of the 2001+ cases that were coded directly) [50]. The following were the values: grade I and II, grade III and IV, unknown (patients' status was either unknown or tumors could not been graded).

Variable 12 – *Radiation* – Patients having radiation as a treatment for their initial lung cancer [50]. The following were the values: yes, no, unknown (status is unknown, or patients refused to receive the treatment).

Variable 13 – *Surgical procedure* – Patients having surgery as a treatment for their lung cancer [50]. The following were the values: yes, no, unknown (status is unknown, or patients refused to go through a surgical procedure).

Variable 14 – *Multiple Primary* – the outcome binary variable that divided the study subjects in two groups: those who developed multiple primaries ("1"), and those who had a single

primary lung cancer ("0"). This variable was used as the "censor" variable for the Cox proportional hazards Regression.

Variable 15 – *Time* - observed survival from the time of the lung cancer diagnosis for those who did not develop another primary cancer, and the difference in observed survival between 1st and 2nd diagnoses for those who developed another primary cancer.

III. RESULTS

3.1. Univariate Descriptive Statistics and Bivariate Findings with Respect to the Exposed and Unexposed Groups. Characteristics of the Study Sample.

In Table 1, basic information on descriptive statistics for lung cancer patients in the exposed and unexposed groups is presented. Means and standard deviations are provided for continuous variables, whereas frequencies and column percentages are provided for categorical variables. For categorical variables, Pearson χ^2 and Fisher's Exact Tests (if the expected cell was $\leq$5) were used to determine if the association between exposure and the outcome was statistically significant, while the Wilcoxon-Mann-Whitney test was used for continuous variables. Two-sided p-values $\leq$.05 were considered statistically significant.

Based on the descriptive statistics from the Table 1 and the results of the Wilcoxon-Mann-Whitney test, the time to development of subsequent cancers was approximately equally distributed in both Appalachian and Non-Appalachian populations (p=0.7147), whereas these two exposure groups were statistically significantly different from each other with respect to distribution of other assessed factors. These include age at diagnosis (p <0.0001), gender (p<0.0001), race (p<0.0001), histology (p<0.0001), tumor stage (p<0.0001), grade (p=0.0002), received treatment of lung cancer, including radiotherapy (p<0.0001) and surgery (p<0.0001), number of months survived after the diagnosis of the lung cancer (p <0.0001), and the number of alive people (p=0.0014).

According to the results presented in Table 1, lung cancer patients who were residents of the Appalachian region, on average, were diagnosed at an earlier age compared to patients from the Non-Appalachian region (65.39 vs. 66.78 years old). Appalachians had more prolonged median survival time from the date of the lung cancer diagnosis (35 vs. 8 months), albeit the

average survival time for patients from Appalachian region was shorter (17.29 vs. 19.37 months). Overall, the Appalachian population had a smaller percentage of those who were still alive at the end of the study period (17.83% vs. 19.48%).

As compared to Non-Appalachians, patients residing in the Appalachian region had a slightly shorter average time to develop subsequent cancers (30.79 vs. 31.13 months). However, this difference was not statistically significant (p=0.7147).

The percentage of men diagnosed with lung cancer among the Appalachian population was higher (62.73% vs. 57.08%), and fewer female subjects were diagnosed with lung cancer when comparing Appalachian and Non-Appalachian populations (37.27% vs. 42.92%).

There were fewer blacks or patients of other races (American Indian/AK Native, Asian/Pacific Islanders) among the Appalachian population (1.29% vs. 8.82%, 0.12% vs. 0.47%, respectively), whereas there were more whites in the Appalachian population compared to the Non-Appalachian population (98.59% vs. 90.71%).

The Appalachian population had a smaller percentage of those with adenocarcinomas (37.08% vs. 44.69%), but larger percentages of those diagnosed and coded as having non-small and squamous cell carcinomas (25.07% vs. 20.75%, 37.85% vs. 34.56%, respectively).

There were smaller percentages of patients who were residents of the Appalachian region, diagnosed with stage II and IV tumors (17.69% vs. 18.51%, 50.28% vs. 51.34%, respectively), albeit the Appalachian population was characterized with a larger percentages of patients with stage III and those with unknown stage tumors (27.17% vs. 26.69%, 4.86% vs. 3.47%). No lung cancer patient had a stage I tumor.

There were fewer patients residing in the Appalachian region that had tumors of grade I and II, as well as those whose tumor grade was either unknown or impossible to be defined (21.31% vs. 23%, 44.34% vs. 44.9%, respectively); whereas there was a larger percentage of the Appalachian population who had grade III and IV tumors (34.35% vs. 32.1%).

Similar trends could also be observed for two treatment variables: radiotherapy and surgical procedure. Appalachian lung cancer residents received radiotherapy and had surgery performed more often (46.62% vs. 45.07%, 67.27% vs. 64.03%, for radiotherapy and surgery respectively), albeit more Appalachians either refused the prescribed treatment or did not survive long enough to receive the treatment (12.55% vs. 9.02%, 7.01% vs. 4.09%).

In addition, there were slight, yet statistically significant, differences between Appalachian and Non-Appalachian patients regarding the overall number of primary cancers they developed during their lifetime ($p=0.0422$). Appalachians had higher percentages of those who had only lung cancer (96.38% vs. 95.79%) and those who had 3 primaries, i.e., those who developed two more primary cancers beside lung cancer (0.45% vs. 0.41%). On the other hand, Non-Appalachians had slightly higher percentages of those who developed one (3.74% vs. 3.15%) and three primaries after being diagnosed with lung cancer (0.07% vs. 0.02%).

Table 1. Descriptive statistics for patients in exposed and unexposed groups. Means and standard deviations (in parentheses) are provided for continuous variables, whereas frequencies and column percentages (in parentheses) are provided for categorical variables. The two-sided p-values $\leq .05$ for Pearson χ^2 and Fisher's Exact, and the Wilcoxon-Mann-Whitney tests were considered statistically significant (N=26456).

Variable	Appalachia (8536, 32.26%)	Non-Appalachia (17920, 67.74%)	P-value
Age at diagnosis (years)			
n	8536	17920	<.0001****
Mean (SD)	65.39 (10.41)	66.78 (10.87)	
Median (Q1, Q3)	66 (58, 73)	67 (59, 75)	
(Min, Max)	(30, 100)	(21, 99)	
# Survival months after diagnosis of Lung Cancer			
n	8528	17892	<.0001****
Mean (SD)	17.29 (25.51)	19.37 (27.74)	
Median (Q1, Q3)	35 (17, 57)	8 (2,23)	
(Min, Max)	(4, 136)	(0, 155)	
Time to Event (months) (development of a subsequent primary cancer)			
n,	309 (3.62)	755 (4.21)	0.7147^{ns}
Mean (SD)	41.65 (30.79)	41.06 (31.13)	
Median (Q1, Q3)	35 (17, 57)	34 (16, 59)	
(Min, Max)	(4, 136)	(4, 154)	

Variable	Appalachia (8536, 32.26%)	Non-Appalachia (17920, 67.74%)	P-value
Gender			
n	8536	17920	<.0001****[2]
Female n, (%)	3181 (37.27)	7691 (42.92)	
Male n, (%)	5355 (62.73)	10229 (57.08)	
Race			
n	8536	17920	<.0001****
Black	110 (1.29)	1580 (8.82)	
White	8416 (98.59)	16256 (90.71)	
Other (American Indian/AK Native, Asian/Pacific Islander)	10 (0.12)	84 (0.47)	

[2] Fisher's Exact Test

Vital Status			
n	8536	17920	.0014***[3]
Alive n, (%)	1522 (17.83)	3491 (19.48)	
Dead n, (%)	7014 (82.17)	14429 (80.52)	
# Number of Primaries			
n	8536	17920	0.0422*
Lung cancer only n, (%)	8227 (96.38)	17165 (95.79)	
LC plus another cancer n, (%)	269 (3.15)	670 (3.74)	
LC plus 2 more cancers n, (%)	38 (0.45)	73 (0.41)	
LC plus 3 more cancers n, (%)	2 (0.02)	12 (0.07)	
Histologic Type ICD-O-3			
n	8536	17920	<.0001****
Adenocarcinoma n, (%)	3165 (37.08)	8008 (44.69)	
Non-small (NOS) n, (%)	2140 (25.07)	3719 (20.75)	
Squamous n, (%)	3231 (37.85)	6193 (34.56)	
Stage			
n	8536 (32.26)	17920 (67.74)	<.0001****
In situ n, (%)	0 (0.00)	0 (0.00)	
Localized n, (%)	1510 (17.69)	3317 (18.51)	
Regional n, (%)	2319 (27.17)	4782 (26.69)	
Distant n, (%)	4292 (50.28)	9200 (51.34)	
Unstaged/Unknown n, (%)	415 (4.86)	621 (3.47)	
Grade			
n	8536	17920	0.0002***
Grade I and II n, (%)	1819 (21.31)	4121 (23.00)	
Grade III and IV n, (%)	2932 (34.35)	5753 (32.10)	
Unknown n, (%)	3785 (44.34)	8046 (44.90)	
Radiation			
n	8536	17920	<.0001****
Yes n, (%)	3618 (42.39)	7949 (44.36)	
No n, (%)	3847 (45.07)	8354 (46.62)	
Unknown n, (%)	1071 (12.55)	1617 (9.02)	
Surgical procedure			
n	8536	17920	<.0001****
Yes n, (%)	2472 (28.96)	5132 (28.64)	
No n, (%)	5466 (64.03)	12055 (67.27)	
Unknown n, (%)	598 (7.01)	733 (4.09)	

[3] Fisher's Exact Test

3.2. Univariate Descriptive Statistics and Bivariate Findings with Respect to the Outcome Groups.

In Table 2, basic information on descriptive statistics for patients in each of the two outcome groups is presented. Means and standard deviations are provided for continuous variables, whereas frequencies and row percentages are provided for categorical variables. For categorical variables, Pearson χ^2 and Fisher's Exact Tests (if the expected cell was ≤ 5) were used to determine if the association between exposure and the outcome was statistically significant, while the Wilcoxon-Mann-Whitney test was used for continuous variables. Two-sided p-values $\leq .05$ were considered statistically significant.

As shown in Table 2, the two outcome groups were significantly different with respect to the following factors: age at diagnosis ($p=0.0005$), number of months survived after the diagnosis of lung cancer ($p<0.0001$), place of residence ($p=0.0228$), histology ($p<0.0001$), tumor stage ($p<0.0001$), grade ($p<0.0001$), received treatment of lung cancer, including radiotherapy ($p<0.0001$) and surgery ($p<0.0001$).

A slightly larger percentage of female lung cancer patients developed multiple primaries (4.29% vs. 3.84%), while a little larger percentage of white subjects developed multiple primaries (4.07%), followed by black patients (3.49%), and those of other racial background (1.06%). Hence, no statistically significant differences were observed regarding patients' gender ($p=0.0699$) and race ($p=0.1661$).

Those who developed subsequent primary cancers after having been diagnosed with lung cancer, were on average one year younger (65.44 vs. 66.37 years old). Patients who developed multiple primaries had a much longer average and median survival time after the date of diagnosis for the initial lung cancer (67.73 vs. 16.81 months, 59 vs. 7 months, respectively).

More patients residing in the Non-Appalachian region developed multiple primaries as compared to Appalachian patients (4.21% vs. 3.62).

Approximately the same percentages of patients developed multiple primary cancers among those who had squamous and adenocarcinomas (4.73%, 4.48%, respectively); however, only two percent of patients with records of non-small cell carcinomas developed multiple primaries.

Comparing lung cancer patients with different tumor stages, we can observe that a larger percentage of people with stage II tumors developed subsequent primary cancers (10.46%), followed by subjects with stage III tumors (5.76%), those whose tumor stage was unknown (3.96), and those with stage IV tumors (0.81%).

Similarly, there were more patients with multiple primary cancers who were diagnosed with grade I and II tumors in the lung (8.01%), followed by those with grade III and IV tumors (4.58%), and 1.61% of patients with ungraded tumors.

Approximately the same percentages of those subjects who did not receive radiotherapy as the treatment for the initial lung cancer (5.20%) and those whose status was unknown (5.13%) eventually developed multiple primaries. However, only roughly half (2.52%) of patients who received radiotherapy developed subsequent cancers.

Conversely, among those patients who had gone through a surgical procedure as the treatment of the initial lung cancer, 10.89 percent developed multiple primary cancers, whereas among those who did not have surgery, only 1.21 percent developed subsequent primary cancers, similarly to the percentage of those who refused the surgery or whose status was unknown (1.8%).

Table 2. Descriptive statistics for patients in two outcome groups. Means and standard deviations (in parentheses) are provided for continuous variables, whereas frequencies and column percentages (in parentheses) are provided for categorical variables. The two-sided p-values $\leq .05$ for Pearson χ^2 and Fisher's Exact, and the Wilcoxon-Mann-Whitney tests were considered statistically significant (N=26456).

Variable	Multiple primaries N=1064	Single primary N=25392	P-value
Age at diagnosis (years)			
N	1064	25392	.0005***
Mean (SD)	65.44 (8.91)	66.37 (10.81)	
Median (Q1, Q3)	66 (59, 72)	67 (59, 74)	
(Min, Max)	(38, 90)	(21, 100)	
# Survival months after 1st diagnosis			
N	1064	25356	<.0001****
Mean (SD)	63.73 (36.68)	16.81 (24.85)	
Median (Q1, Q3)	59 (32, 89)	7 (2, 20)	
(Min, Max)	(4, 155)	(0, 155)	

N	Variable	Multiple primaries N=1064	Single primary N=25392	P-value
	Appalachia region			
26456	n	1064	25392	
8536	Yes n, (row %)	309 (3.62)	8227 (96.38)	0.0228*[4]
17920	No n, (row %)	755 (4.21)	17165 (95.79)	
	Gender			
26456	n	1064	25392	0.0699 ns[5]
19872	Female n, (row %)	466 (4.29)	10406 (95.71)	
15584	Male n, (row %)	598 (3.84)	14986 (96.16)	
	Race			
26456	n	1064	25392	0.1661 ns
1690	Black n, (row %)	59 (3.49)	1631 (96.51)	
24672	White n, (row %)	1004 (4.07)	23668 (95.93)	
94	Other (American Indian/AK Native, Asian/Pacific Islander/Unknown) n, (row %)	1 (1.06)	93 (98.94)	
	Histologic Type ICD-O-3			
26456	n	1064	26356	

[4] Fisher's Exact Test
[5] Fisher's Exact Test

11173	Adenocarcinoma n, (row %)	501 (4.48)	10672 (95.52)	<.0001***
5859	Non-small (NOS) n, (row %)	117 (2.00)	5742 (98.00)	*
9424	Squamous n, (row %)	446 (4.73)	8978 (95.27)	
	Stage			
26456	n	1064	26356	
0	In situ n, (row %)	0 (0.00)	0 (0.00)	
4827	Localized n, (row %)	505 (10.46)	4322 (89.54)	<.0001***
7101	Regional n, (row %)	409 (5.76)	6692 (94.24)	*
13492	Distant n, (row %)	109 (0.81)	13383 (99.19)	
1036	Unknown n, (row %)	41 (3.96)	995 (96.04)	
	Grade			
26456	n	1064	26356	
5940	Grade I and II n, (row %)	476 (8.01)	5464 (91.99)	<.0001***
8685	Grade III and IV n, (row %)	398 (4.58)	8287 (95.42)	*
11831	Unknown n, (row %)	190 (1.61)	11641 (98.39)	
	Radiation			
26456	n	1064	26356	
11567	Yes n, (row %)	291 (2.52)	11276 (97.48)	<.0001***
12201	No n, (row %)	635 (5.20)	11566 (94.80)	*
2688	Unknown n, (row %)	138 (5.13)	2550 (94.87)	
	Surgical procedure			
26456	n	1064	26356	
7604	Yes n, (row %)	828 (10.89)	6776 (89.11)	<.0001***
17521	No n, (row %)	212 (1.21)	17309 (98.79)	*
1331	Unknown n, (row %)	24 (1.80)	1307 . 20)	

3.3. Bivariate Findings.

Table 3 presents the results from univariate analysis using Cox proportional hazards regression models, with development of multiple primary lung cancers as the “censor” variable, the “time” variable which expressed observed survival from the time of the lung cancer diagnosis for those who did not develop another primary cancer, and the difference in observed survival between 1st and 2nd diagnoses for those who developed another primary cancer. Univariate analysis allowed assessing unadjusted associations.

The univariate Cox regression models did not reveal statistically significant association for the primary exposure – residing in Appalachian region (HR = 0.993, 95% CI: 0.869-1.133).

The model also did not demonstrate significant results for the race covariate. The observed results were the following: whites had 12.1 percent higher observed hazards to develop multiple primary cancers as compared to blacks (95% CI: 0.862-1.457), whereas patients of other nationalities had 67.3 percent smaller hazards comparing to blacks (95% CI: 0.046-2.348).

On the other hand, univariate Cox proportional hazards regressions revealed a number of other statistically significant associations. For instance, being diagnosed with lung cancer at a later age increased the hazards of development of a subsequent primary cancer by 1.5 percent per increase in one year (95% CI: 1.008-1.021).

Males had 16.5 percent higher hazards to develop subsequent cancers as compared to female lung cancer patients (95% CI: 1.032-1.315).

Furthermore, a statistically significant protective effect of being diagnosed with a non-small lung carcinoma was observed, decreasing the hazards of developing a subsequent primary cancer by 31.8 percent (95% CI: 0.556-0.836). Patients who had adenocarcinomas had 21.4 percent smaller hazards to develop another primary cancer as compared to squamous cell carcinomas. However, the former results did not have sufficient statistical significance (95% CI: 0.780-1.007).

A statistically non-significant association was observed for those with unknown status for radiotherapy (HR=0.84, 95% CI: 0.697-1.011). However, compared to those who did not receive radiation as the treatment for lung cancer, patients who had radiation had 18.9 percent smaller hazards to develop multiple primaries (95% CI: 0.705-0.933).

Compared to those who did not have surgery as a treatment for the initial lung cancer, patients who had a surgical procedure performed had 1.7 times (95% CI: 1.452-1.991) higher

hazards to develop a subsequent cancer, whereas those with unknown status had 5.3 percent smaller hazards, albeit the association was non-significant (95% CI: 0.621-1.446).

Patients who had grade III and IV tumors had 13.3 percent smaller hazards as compared to those with grade I and II tumors (95% CI: 0.759-0.991). Study subjects whose status was unknown or whose tumors were ungraded had 40.5 percent smaller hazards as compared to those with grade I and II tumors. (95% CI: 0.502-0.706).

The following results also demonstrated that having a stage IV tumor decreased the hazards of the outcome by 48.6% comparing to the patients with stage II tumors (95% CI: 0.415-0.636). Having stage III tumors also had an observed protective effect from developing multiple primary cancers, yet the results were not significant (HR=0.897, 95% CI: 0.787-1.022). Those patients whose status was unknown had 24.6 percent higher hazards to develop multiple primary cancers, albeit that association was not statistically significant either (95% CI: 0.548-1.036).

Table 3. Univariate Cox proportional hazards models' results, using development of multiple primary lung cancers as the "censor" variable, and the "time" variable which expressed observed survival from the time of the lung cancer diagnosis for those who did not develop another primary cancer, and the difference in observed survival between 1st and 2nd diagnoses for those who developed another primary cancer (N=26456).

Variable	HR	95% CI	P-value
Age at diagnosis (years)	1.014	1.008- 1.020	<0.0001
Appalachia region			
No*			
Yes	0.993	0.869- 1.133	0.9117
Gender			
Female*			
Male	1.165	1.032-1.315	0.0138
Race			
Black*			
White	1.121	0.862- 1.457	0.3946
Other	0.327	0.046- 2.348	0.2667
Histologic Type ICD-O-3			
Squamous*			
Adenocarcinoma	0.886	0.780- 1.007	0.0642
Non-Small (NOS)	0.682	0.556- 0.836	0.0002
Radiotherapy			
No*			
Yes	0.811	0.705-0.933	0.0035
Unknown	0.840	0.697-1.011	0.0651
Surgical procedure			
No			
Yes*	1.700	1.452-1.991	<0.0001
Unknown	0.947	0.621-1.446	0.8025
Grade			
Grade I and II *			
Grade III and IV	0.867	0.759-0.991	0.0365
Unknown	0.595	0.502-0.706	<0.0001
Tumor Stage			
Localized*			
Regional	0.897	0.787- 1.022	0.1022
Distant	0.514	0.415- 0.636	<0.0001
Unknown	0.754	0.548- 1.036	0.0819

HR – hazard ratio, CI – confidence interval

*reference group

3.4. Multivariable Findings.

Table 4 presents results from the final Cox proportional hazards model, using development of multiple primary cancers as the "censor" variable, the "time" variable which expressed observed survival from the time of the lung cancer diagnosis for those who did not develop another primary cancer, and the difference in observed survival between 1st and 2nd diagnoses for those who developed another primary cancer (N=26456).

The final model indicated that there were no statistically significant differences between Appalachian and Non-Appalachian populations with respect to the hazards of developing a subsequent primary cancer (HR: 1.002, 95% CI: 0.877-1.145).

The adjusted analysis revealed that increasing age at diagnosis by one year was associated with increase in hazards of development of multiple primary cancer by 1.5 percent (95% CI:1.008-1.021).

Males had 16.9 percent higher hazards to develop subsequent cancers compared to female lung cancer patients (95% CI: 1.035-1.320).

Comparing to patients who had not had a surgery as a treatment of the initial lung cancer, to patients who had surgical procedure performed have shown a 1.446 times (95% CI: 1.183-1.767) higher hazards to develop a subsequent cancer, whereas those with unknown status had 8.6% smaller hazards, albeit the association was not significant (95% CI: 0.583-1.433).

A statistically non-significant association was also observed for radiotherapy. As compared to those who did not receive radiation as their treatment for lung cancer, patients who had radiation therapy had 9.6 percent higher hazards of developing multiple primaries (95% CI:

0.932-1.290), and those with unknown status had 7.6 percent smaller hazards (95% CI: 0.764-1.117).

Patients who had grade III and IV tumors had 7.8 percent smaller hazards as compared to those with grade I and II tumors. However, the association was not statistically significant (95% CI: 0.805-1.056). Study subjects whose status was unknown or whose tumors were ungraded had 23 percent smaller hazards as compared to those with grade I and II tumors (95% CI: 0.634-0.937).

The following results demonstrated that having a stage IV tumor decreased the hazards of the outcome by 31.6% comparing to the patients with stage II tumors (HR=0.684, 95% CI: 0.540-0.865). Having stage III tumors also had an observed protective effect from developing multiple primary cancers, however, the results were not significant (HR= 0.942, 95% CI: 0.821-1.082). Those patients whose status was unknown had 1.7 percent higher hazards of developing of multiple primary cancers, albeit that association was not statistically significant either (95% CI: 0.723-1.433).

Table 4. Multivariable Cox proportional hazards regression's results for the final model of analysis of factors affecting development of multiple primary cancers as the "censor" variable, the "time" variable which expressed observed survival from the time of the lung cancer diagnosis for those who did not develop another primary cancer, and the difference in observed survival between 1st and 2nd diagnoses for those who developed another primary cancer (N=26456)

Variable	Adjusted HR	95% CI	P-value
Appalachia region			
No*			
Yes	1.002	0.877-1.145	0.9713
Age at diagnosis (years)	1.015	1.008-1.021	0.0001
Gender			
Female*			
Male	1.169	1.035-1.320	0.0120
Surgical procedure			
No*			
Yes	1.446	1.183-1.767	0 .0003
Unknown	.914	0.583-1.433	0.6965
Radiotherapy			
No*			
Yes	1.096	0.932-1.290	0.2677
Unknown	0.924	0.764-1.117	0.4142
Grade			
Grade I and II *			
Grade III and IV	0.922	0.805-1.056	0.2423
Unknown	0.770	0.634-0.937	0.0089
Tumor Stage			
Localized*			
Regional	0.942	0.821-1.082	0.4004
Distant	0.684	0.540-0.865	0.0015
Unknown	1.017	0.723-1.433	0.9210

HR – hazard ratio, CI – confidence interval

*reference group

IV. DISCUSSION

4.1. Interpretation of the Results. Consistency with Previous Studies.

The aim of this study was to examine whether there were differences in the development of subsequent primary cancers in lung cancer patients residing in the Appalachian versus Non-Appalachian regions of Kentucky. Interestingly, based on the descriptive statistics and bivariate results presented in Table 1, Appalachian and Non-Appalachian lung cancer patients were different with respect to a number of factors. Namely, these include age at diagnosis of lung cancer, survival months, gender, race, histology, stage and grade of tumors, and received treatment (radiation and surgery). Appalachian and Non-Appalachian residents were significantly different even based on the number of primary cancers which a patient could develop during the lifetime.

However, these two exposure groups did not differ significantly with respect to the observed mean and median time of development of a subsequent primary cancer. Thus, both univariate and multivariable hazards models did not reveal any differences between Appalachian and Non-Appalachian lung cancer patients with respect to the outcome. It was concluded that, even though the variable "Number of Primary Cancers" was treated as a categorical one, statistical significance of a Pearson χ^2 presented in Table 1, could be still explained by a large sample size, and thus, the power of the test.

Overall, the multivariable proportional hazards model revealed the results that were consistent with previous research, or that could be intuitively derived based on descriptive and bivariate findings. For instance, it was expected to see that patient diagnosed at older age would have higher hazards of developing another primary cancer, as typically the increase in incidence rates of many cancers is associated with increase in age [7].

In addition, other research showed that a more aggressive invasive treatment increased the overall survival for patients with multiple lung cancers [25, 28]. That is why we assumed that surgery was very likely to be associated with getting a second primary, as patients who had surgery were likely to live longer, and thus had a greater opportunity to develop a second primary. In contrast, patients who were diagnosed with stage IV lung cancer had very short survival times and were, thus, less likely to develop a second primary cancer. Moreover, those patients with unknown grade status also had smaller hazards of developing another primary cancer. This could be explained by the fact that in a number of patients it was impossible to differentiate and define the grade since they had stage IV tumors, the factor due to which their overall survival time could have been shortened.

Radiation was expected to be a statistically significant factor, which should increase the hazards of development of second primary cancers. Performing radiation indeed increased the hazards; however, this association was not significant, even though the investigators did not distinguish different types or doses of radiation that patients received, but had rather defined three main groups (radiation performed, not performed, or unknown) in order not to lose power on every additional level.

The results of this study were consistent with previous analysis based on SEER database, which showed that four percent of the patients who had the first primary cancer of the lung or other respiratory organs developed subsequent primary cancers. The median age at diagnosis of lung or other respiratory organs' cancers was 67 years among those patients who developed multiple primary cancers [7]. This study has also revealed that relatively small percentage (4.94%) of lung cancer patients developed subsequent tumor events, and that the median age at diagnosis of patients who developed multiple primary cancers was 67 years. In addition,

consistent with previous studies [7, 41], the results of this study demonstrated that patients had to live long enough to develop subsequent primary cancers. According to Table 2, the mean observed survival time, measured in months, of those who had developed multiple primary cancers was almost four times longer than of those who had single primary lung cancers. Moreover, median survival time was more than eight times longer for the outcome group. Thus, if lung cancer was detected at an earlier stage and treatment could have been prescribed to the patients, thereby increasing their overall survival, it also increased the hazards to develop a second primary cancer. Therefore, very careful attention should be paid to lung cancer patients who demonstrate long overall survival, and especially those who undergo invasive treatment. Additional screening tests should be required to control the condition of this population of lung cancer patients.

Previous studies have shown that female lung cancer patients, especially those with stage II tumors, had better survival compared to men [7]. Some studies explained this phenomenon by the fact that surgical procedures were performed on women with stage II tumors more frequently [7, 42]. Another possible explanation is that women are more likely to develop adenocarcinomas compared to men [7, 52], as Fu et al. showed that patients with any histologic type of stage II tumors other than adenocarcinomas had lower survival rates [42]. In addition, there could be a healthy screening effect as lung cancer could be detected on earlier stages among women who undergo their annual mammograms. Unexpectedly, the results of this study demonstrated that men had almost a 17 percent higher hazards of developing subsequent cancers compared to female subjects. A possible explanation is the inability to control for smoking in this analysis as because smoking is a major risk factor for many cancers, lung cancer in particular, and also for development of a second primary cancer [6, 7]. In addition, it is well known that the male

population of Kentucky has a higher prevalence of smoking [53, 54] compared to women. Therefore, further research is recommended to examine the influence of smoking as a potential confounding factor.

4.2. Strengths and Limitations of the Study.

It should be emphasized that, besides examining whether there were differences in the development of subsequent primary cancers in lung cancer patients residing in the Appalachian versus Non-Appalachian regions of Kentucky, this study also identified other factors associated with the development of another primary cancer in lung cancer patients. Namely, age at diagnosis, gender, race, stage and grade of the tumor, and treatment (radiation or surgery). The final Cox proportional hazards model included all covariates that were considered as potential confounders.

Because a prospective cohort study due to lack of individual-based information, the chosen design of a retrospective, population-based cohort study was the best option. A case-control study design would not be appropriate here, as it would have required a defined cutoff time during which a patient could have developed a subsequent primary cancer, since it would have been critical for investigators to identify “cases” and “controls” correctly. This, in turn, would automatically reduce the sample size, and decrease the power.

The Cox proportional hazards model, which allowed for controlling for the time from diagnosis to death or a subsequent primary cancer was the appropriate model. Adjusted logistic regression would not have allowed for controlling for the time from diagnosis to death or a subsequent primary cancer as well as for considering censoring cases of death due to various causes.

On the other hand, this study had a number of limitations. Firstly, it lacks generalizability. Kentucky lung cancer patients were considered the target population; thus, the results can be applied only to the population of Kentucky lung cancer patients.

Another limitation of this study was that a large volume of missing data was included in order not to lose power in case subjects with missing values for important covariates had been excluded. Further research is recommended to examine how the estimates might change if missing cases were excluded or if the information could be obtained by linking SEER data to other databases (e.g., Medicaid or Medicare).

Unfortunately, there was no individual-level information about smoking status of the patients. The investigators did not distinguish the types of subsequent primary cancers. Some of those could have been subsequent primary lung cancers. Studies exploring predictors of synchronous multiple lung cancers demonstrated that smoking (which is more prevalent among males) is a substantial risk factor [6, 7, 27]. Further research is recommended to examine the influence of smoking as a potential confounding factor.

Moreover, the impact of comorbidities could not be assessed in this study. The SEER database includes variable that defines causes of death; however, the investigators could not use it since a large number of subjects were still alive at the time of defined cutoff. Further research is recommended to address this question.

V. SUMMARY, CONCLUSIONS, RECOMMENDATIONS

This was a retrospective, population-based cohort study of Kentucky patients diagnosed with primary lung cancer between 2000 and 2013. The study population was drawn from the Kentucky Cancer Registry. Subjects were followed to determine if they developed subsequent primary cancers. The Cox proportional hazards model was used to control for the time from diagnosis to death or a second primary cancer.

Even though unadjusted associations revealed some differences between Appalachian and Non-Appalachian populations of Kentucky lung cancer patients, the adjusted multivariable Cox proportional hazards regression analysis indicated that there were no differences between Appalachian and Non-Appalachian lung cancer patients with regards to development of subsequent primary cancers.

Increasing age at diagnosis, as well as male gender and surgical treatment, increased the hazards of developing subsequent cancers, whereas having stage IV or ungraded tumors decreased the hazards of the developing subsequent primary cancers.

The results of this study suggest that a number of improvements in Kentucky's healthcare system with regards to prevention of multiple primary cancers in lung cancer patients can be considered. For instance, clinical surveillance, including early detection of lung cancer, screening of the male population and smokers, making careful and thoughtful choice of treatment, i.e. choosing between radiotherapy and surgery, and also assessment of long-term treatment effects.

Further research is recommended to define whether the following factors could have made a difference to the adjusted estimates: educational attainment, poverty level, employment,

smoking status, number and sequence of treatments (i.e. radiotherapy and surgery), comorbidities, and occupational exposures.

VI. REFERENCES

1. United Health Foundation, *America's Health Rankings.* 2016.
2. Kentucky Department for Public Health, *State Health Assessment.* 2013: p. 9, 54-72.
3. Centers for Disease Control and Prevention, *Cancer Rates by U.S. State.* 2014.
4. Registry, K.C., *Age-Adjusted Cancer Incidence Rates in Kentucky by Appalachian Region.* 2016.
5. Registry, K.C., *Age-Adjusted Mortality Rates in Kentucky by Appalachian Region.* 2016.
6. Koubkova, L., et al., *Second Primary Cancers–Causes, Incidence and the Future.* Klin Onkol, 2014. **27**(1): p. 11-17.
7. Hayat, M.J., et al., *Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program.* The oncologist, 2007. **12**(1): p. 20-37.
8. Bhaskarla, A., et al., *Analysis of Second Primary Lung Cancers in the SEER Database.* Journal of Surgical Research. **162**(1): p. 1-6.
9. Nielsen, S.F., B.G. Nordestgaard, and S.E. Bojesen, *Associations between first and second primary cancers: a population-based study.* Canadian Medical Association Journal, 2011: p. cmaj. 110167.
10. Curtis, R.E., et al., *New malignancies among cancer survivors.* SEER cancer registries, 1973. **2000**.
11. Chakraborty, S., et al., *Increased incidence of a second lymphoproliferative malignancy in patients with multiple myeloma–a SEER based study.* Anticancer research, 2012. **32**(10): p. 4507-4515.
12. Chakraborty, S., et al., *Incidence and Prognostic Significance of Second Primary Cancers in Renal Cell Carcinoma.* American journal of clinical oncology, 2013. **36**(2): p. 132-142.
13. Nsouli-Maktabi, H.H., et al., *Second primary breast, endometrial, and ovarian cancers in Black and White breast cancer survivors over a 35-year time span: effect of age.* Breast Cancer Research and Treatment, 2011. **129**(3): p. 963-969.
14. American Cancer Society, *Lung Cancer Key Statistics.* 2016.
15. Centers for Disease Control and Prevention, *Lung Cancer.* 2016. **Cancer**.
16. American Cancer Society, *Cancer Facts & Figures 2015.* 2015: p. 7-8.
17. Kentucky Cancer Registry, *Cancer of Lung and Bronchus, 2000-2013.* Age-Adjusted Invasive Cancer Incidence Rates in Kentucky, 2013.
18. Kentucky Cancer Registry, *Cancer of Lung and Bronchus, 2000-2012.* Age-Adjusted Cancer Mortality Rates in Kentucky, 2012.
19. Khuri, F.R., et al., *The impact of smoking status, disease stage, and index tumor site on second primary tumor incidence and tumor recurrence in the head and neck retinoid chemoprevention trial.* Cancer Epidemiology Biomarkers & Prevention, 2001. **10**(8): p. 823-829.
20. Begg, C.B., et al., *Methodology for Evaluating the Incidence of Second Primary Cancers with Application to Smoking-relted Cancers from the Surveillance, Epidmiology, and End Results (SEER) Program.* American journal of epidemiology, 1995. **142**(6): p. 653-665.

21. Bi, R., et al., *Malignant ameloblastoma (metastatic ameloblastoma) in the lung: 3 cases of misdiagnosis as primary lung tumor with a unique growth pattern.* Diagn Pathol, 2015. **10**: p. 123.
22. Xue, X., et al., *Diagnosis of multiple primary lung cancer: A systematic review.* Journal of International Medical Research, 2013. **41**(6): p. 1779-1787.
23. Homer, R.J., *Pathologists' staging of multiple foci of lung cancer: poor concordance in absence of dramatic histologic or molecular differences.* Am J Clin Pathol, 2015. **143**(5): p. 701-6.
24. Murphy, S.J., et al., *Identification of independent primary tumors and intrapulmonary metastases using DNA rearrangements in non-small-cell lung cancer.* Journal of Clinical Oncology, 2014: p. JCO. 2014.56. 7644.
25. Vansteenkiste, J.F., et al., *Practical approach to patients presenting with multiple synchronous suspect lung lesions: a reflection on the current TNM classification based on 54 cases with complete follow-up.* Lung Cancer, 2001. **34**(2): p. 169-175.
26. Zhang, Y., et al., *Synchronous non-small cell lung cancers: diagnostic yield can be improved by histologic and genetic methods.* Ann Surg Oncol, 2014. **21**(13): p. 4369-74.
27. Mizushima, Y. and M. Kobayashi, *Clinical characteristics of synchronous multiple lung cancer associated with idiopathic pulmonary fibrosis: a review of Japanese cases.* CHEST Journal, 1995. **108**(5): p. 1272-1277.
28. Nakata, M., et al., *Surgical treatments for multiple primary adenocarcinoma of the lung.* The Annals of thoracic surgery, 2004. **78**(4): p. 1194-1199.
29. Martini, N. and M. Melamed, *Multiple primary lung cancers.* The Journal of Thoracic and Cardiovascular Surgery, 1975. **70**(4): p. 606-612.
30. Trousse, D., et al., *Synchronous multiple primary lung cancer: an increasing clinical occurrence requiring multidisciplinary management.* The Journal of Thoracic and Cardiovascular Surgery, 2007. **133**(5): p. 1193-1200.
31. Adebonojo, S.A., D.M. Moritz, and C.A. Danby, *The results of modern surgical therapy for multiple primary lung cancers.* CHEST Journal, 1997. **112**(3): p. 693-701.
32. Okada, M., et al., *Operative approach for multiple primary lung carcinomas.* The Journal of thoracic and cardiovascular surgery, 1998. **115**(4): p. 836-840.
33. Girard, N., et al., *Comprehensive histologic assessment helps to differentiate multiple lung primary nonsmall cell carcinomas from metastases.* The American journal of surgical pathology, 2009. **33**(12): p. 1752-1764.
34. Berg, J.W., D. Schottenfeld, and F. Ritter, *Incidence of multiple primary cancers. III. Cancers of the respiratory and upper digestive system as multiple primary cancers.* Journal of the National Cancer Institute, 1970. **44**(2): p. 263-274.
35. Wu, S.-c., et al., *Multiple primary lung cancers.* CHEST Journal, 1987. **92**(5): p. 892-896.
36. Chang, Y.-L., et al., *Clonality and prognostic implications of p53 and epidermal growth factor receptor somatic aberrations in multiple primary lung cancers.* Clinical Cancer Research, 2007. **13**(1): p. 52-58.
37. Rosengart, T.K., et al., *Multiple primary lung carcinomas: prognosis and treatment.* The Annals of thoracic surgery, 1991. **52**(4): p. 773-779.

38. Sozzi, G., et al., *Genetic evidence for an independent origin of multiple preneoplastic and neoplastic lung lesions.* Cancer research, 1995. **55**(1): p. 135-140.
39. Mesmoudi, M., et al., *A Synchronous undifferentiated nasopharyngeal carcinoma and infiltrating ductal carcinoma of the breast successfully treated with induction chemotherapy followed by local control of both tumours: a case report.* BMC Ear, Nose and Throat Disorders, 2011. **11**(1): p. 6.
40. Sawka, A.M., et al., *Second primary malignancy risk after radioactive iodine treatment for thyroid cancer: a systematic review and meta-analysis.* Thyroid, 2009. **19**(5): p. 451-457.
41. Martini, N., et al., *Incidence of local recurrence and second primary tumors in resected stage I lung cancer.* The Journal of Thoracic and Cardiovascular Surgery, 1995. **109**(1): p. 120-129.
42. Fu, J.B., et al., *Lung cancer in women: analysis of the national Surveillance, Epidemiology, and End Results database.* CHEST Journal, 2005. **127**(3): p. 768-777.
43. Rodrigues, G., et al., *Definitive radiation therapy in locally advanced non-small cell lung cancer: Executive summary of an American Society for Radiation Oncology (ASTRO) evidence-based clinical practice guideline.* Practical radiation oncology, 2015. **5**(3): p. 141-148.
44. Geryk, E., et al., *[Cancer multiplicities in the Czech population].* Casopis lekaru ceskych, 2009. **149**(4): p. 178-183.
45. Schoenberg, N.E., et al., *Trends in cigarette smoking and obesity in Appalachian Kentucky.* South Med J, 2015. **108**(3): p. 170-7.
46. ATRN, A.T.R.N., *Appalachia & Appalachian Health: Quick Facts.*
47. Surveillance, E., and End Results Program,, *County Attributes.* 2016.
48. American Cancer Society, *Probability (%) of Developing Invasive Cancer during Selected Age Intervals by Sex, US, 2009-2011.* 2015: p. 14.
49. Institute, N.C., *Surveillance, Epidemiology, and End Results Program.*
50. Surveillance, E., and End Results Program,, *Dictionary of SEER*Stat Variables* November 2011 Submission (released April 2012), 2012.
51. Lewis, D.R., et al., *US lung cancer trends by histologic type.* Cancer, 2014. **120**(18): p. 2883-2892.
52. Devesa, S.S., et al., *International lung cancer trends by histologic type: Male:Female differences diminishing and adenocarcinoma rates rising.* International Journal of Cancer, 2005. **117**(2): p. 294-299.
53. Centers for Disease Control and Prevention, *State-Specific Prevalence of Cigarette Smoking and Smokeless Tobacco Use Among Adults --- United States, 2009.* 2010.
54. Kentucky Department for Public Health, *State Health Assessment.* 2013: p. 25-26.

APPENDIX

Table 1. Model 1: All covariates except Race.

Analysis of Maximum Likelihood Estimates

Parameter		*DF*	*Parameter Estimate*	*Standard Error*	*Chi-Square*	*Pr > ChiSq*	*Hazard Ratio*	*95% Hazard Ratio Confidence Limits*	
Appalachia_Region__A	*Appalachia*	1	-0.0002094	0.06800	0.0000	0.9975	1.000	0.875	1.142
Age_at_diagnosis1		1	0.01405	0.00320	19.2435	<.0001	1.014	1.008	1.021
Histology1	*Adenocarcinoma*	1	-0.07646	0.06725	1.2924	0.2556	0.926	0.812	1.057
Histology1	*Non–small cell carcinoma*	1	-0.14585	0.10869	1.8005	0.1797	0.864	0.698	1.069
GradeFirst	*Grade III and IV*	1	-0.07528	0.07069	1.1340	0.2869	0.927	0.807	1.065
GradeFirst	*Unknown*	1	-0.24055	0.10164	5.6017	0.0179	0.786	0.644	0.959
*Summary_stage_2000_*	*Distant*	1	-0.37506	0.12014	9.7461	0.0018	0.687	0.543	0.870
*Summary_stage_2000_*	*Regional*	1	-0.06356	0.07065	0.8095	0.3683	0.938	0.817	1.078
*Summary_stage_2000_*	*Unstaged/Unknown*	1	0.01644	0.17472	0.0089	0.9250	1.017	0.722	1.432
Sex1	*Male*	1	0.14592	0.06291	5.3805	0.0204	1.157	1.023	1.309
Radiotherapy1	*Radiation performed*	1	0.08765	0.08306	1.1136	0.2913	1.092	0.928	1.285
Radiotherapy1	*Unknown*	1	-0.08356	0.09718	0.7394	0.3899	0.920	0.760	1.113
Surgery1	*Surgical procedure performed*	1	0.36172	0.10322	12.2795	0.0005	1.436	1.173	1.758
Surgery1	*Unknown*	1	-0.09118	0.22946	0.1579	0.6911	0.913	0.582	1.431

Stratification by YEAR of DIAGNOSIS

Table 2. Group 2000-2006, Model 2.

Analysis of Maximum Likelihood Estimates

Parameter		*DF*	*Parameter Estimate*	*Standard Error*	*Chi-Square*	*Pr > ChiSq*	*Hazard Ratio*	*95% Hazard Ratio Confidence Limits*	
Appalachia_Region__A	*Appalachia*	1	0.03624	0.08223	0.1942	0.6594	1.037	0.883	1.218
Age_at_diagnosis1		1	0.01377	0.00388	12.6265	0.0004	1.014	1.006	1.022
GradeFirst	*Grade III and IV*	1	-0.04737	0.08206	0.3332	0.5638	0.954	0.812	1.120
GradeFirst	*Unknown*	1	-0.45484	0.12636	12.9561	0.0003	0.635	0.495	0.813
*Summary_stage_2000_*	*Distant*	1	-0.28608	0.15756	3.2970	0.0694	0.751	0.552	1.023
*Summary_stage_2000_*	*Regional*	1	-0.03754	0.08577	0.1916	0.6616	0.963	0.814	1.139
*Summary_stage_2000_*	*Unstaged/Unknown*	1	0.14588	0.18514	0.6209	0.4307	1.157	0.805	1.663
Sex1	*Male*	1	0.09888	0.07509	1.7340	0.1879	1.104	0.953	1.279
Radiotherapy1	*Radiation performed*	1	-0.04066	0.10448	0.1515	0.6971	0.960	0.782	1.178
Radiotherapy1	*Unknown*	1	-0.04087	0.10441	0.1532	0.6955	0.960	0.782	1.178
Surgery1	*Surgical procedure performed*	1	0.34516	0.13183	6.8546	0.0088	1.412	1.091	1.829
Surgery1	*Unknown*	1	0.00286	0.25633	0.0001	0.9911	1.003	0.607	1.657

Table 3. Group 2007-2013, Model 2.

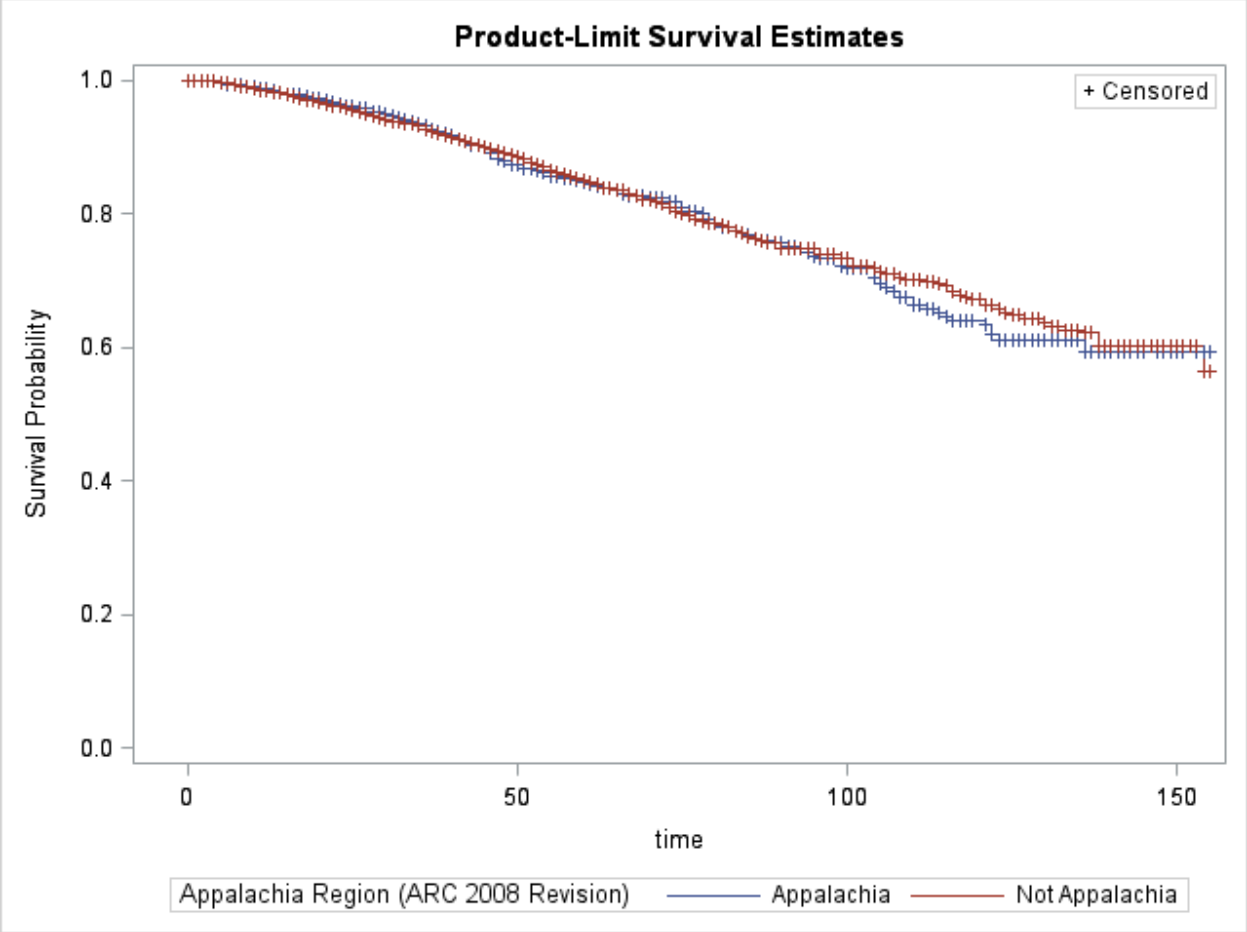
Analysis of Maximum Likelihood Estimates

Parameter		*DF*	*Parameter Estimate*	*Standard Error*	*Chi-Square*	*Pr > ChiSq*	*Hazard Ratio*	*95% Hazard Ratio Confidence Limits*	
Appalachia_Region__A	*Appalachia*	1	-0.08837	0.12093	0.5341	0.4649	0.915	0.722	1.160
Age_at_diagnosis1		1	0.01611	0.00557	8.3570	0.0038	1.016	1.005	1.027
GradeFirst	*Grade III and IV*	1	-0.13513	0.13078	1.0677	0.3015	0.874	0.676	1.129
GradeFirst	*Unknown*	1	0.12842	0.17426	0.5431	0.4611	1.137	0.808	1.600
*Summary_stage_2000_*	*Distant*	1	-0.51394	0.18722	7.5356	0.0060	0.598	0.414	0.863
*Summary_stage_2000_*	*Regional*	1	-0.08307	0.12476	0.4433	0.5055	0.920	0.721	1.175
*Summary_stage_2000_*	*Unstaged/Unknown*	1	-0.79842	0.71965	1.2309	0.2672	0.450	0.110	1.844
Sex1	*Male*	1	0.29771	0.11123	7.1636	0.0074	1.347	1.083	1.675
Radiotherapy1	*Radiation performed*	1	0.32725	0.14193	5.3161	0.0211	1.387	1.050	1.832
Radiotherapy1	*Unknown*	1	-0.36979	0.31602	1.3692	0.2419	0.691	0.372	1.284
Surgery1	*Surgical procedure performed*	1	0.57430	0.17415	10.8749	0.0010	1.776	1.262	2.498
Surgery1	*Unknown*	1	-0.44879	0.59688	0.5653	0.4521	0.638	0.198	2.057

Figure 1. K-M Curve: There are no statistically significant differences between the exposed and unexposed groups not controlling for confounders

Test of Equality over Strata

Test	Chi-Square	DF	Pr >
Chi-Square Log-Rank	0.0124	1	0.9115
Wilcoxon	0.6657	1	0.4146
-2Log(LR)	0.4567	1	0.4992



According to the Figure 1, the log-rank tests revealed that there were no statistically significant differences between Appalachian and Non-Appalachian populations with respect to the probability of developing subsequent primary cancers over 155 months (59.44% vs. 56.5%, respectively; log-rank test: $\chi 2=0.0124$, $p=.9115$).

The study cohort included patients who were diagnosed from January 1, 2000 through December 31, 2013. Assuming that the medical care capabilities, including detection, diagnosing, and treatment of lung cancer might have changed over the thirteen-year time period, the investigators made an attempt of stratification based on the years of diagnosis of the initial lung cancer. Subjects were divided into two subgroups: those who were diagnosed between 2000 and 2006 (n=13,611), and those diagnosed between 2007 and 2013 (n=12,845). This stratification was applied to carry out the Multiple Cox proportional hazards Regression to examine if the hazards would change. In addition, such a stratification was applied to determine if the Kaplan-Meier estimated survival curves would be visually modified. For those patients diagnosed with lung cancer between 2000 and 2006, the multivariable proportional hazards model revealed that only age at diagnosis, performing surgery and having an ungraded tumor were statistically significant predictors (95% CIs: 1.006-1.022, 1.091-1.829, and 0.495-0.813, respectively). On the other hand, for those patients diagnosed with lung cancer between 2007 and 2013, the multivariable proportional hazards model demonstrated that age at diagnosis, gender, having a stage IV tumor, radiation and surgery were statistically significant predictors (95% CIs: 1.005-1.027, 1.083-1.675, 0.414-0.863, 1.05-1.832, 1.262-2.498, respectively)[6]. Thus, the investigators concluded that statistically significant associations which were observed in the final multivariable proportional hazards model for non-stratified data could have been influenced by patients diagnosed in earlier years compared to those diagnosed in later years. However, overall, the stratification did not make a significant difference to the estimates of the multivariable proportional hazards model as well as to the Kaplan-Meier estimated survival curves (See Figure

[6] Refer to Table 1 and Table 2 in the Appendix

1.1. and 1.2.). Thus, it was decided to report the results for the original, non-stratified data in order to increase the power with a larger sample.

Figure 1.1. The K-M Estimated Survival Curve ≤2006 (this stratum includes patients who were diagnosed with lung cancer in 2000 through 2006).

There are no statistically significant differences between the exposed and unexposed groups not controlling for confounders

Test of Equality over Strata

Test	Chi-Square	DF	Pr >
Chi-Square Log-Rank	0.1208	1	0.7282
Wilcoxon	0.0389	1	0.8437
-2Log(LR)	0.0100	1	0.9205

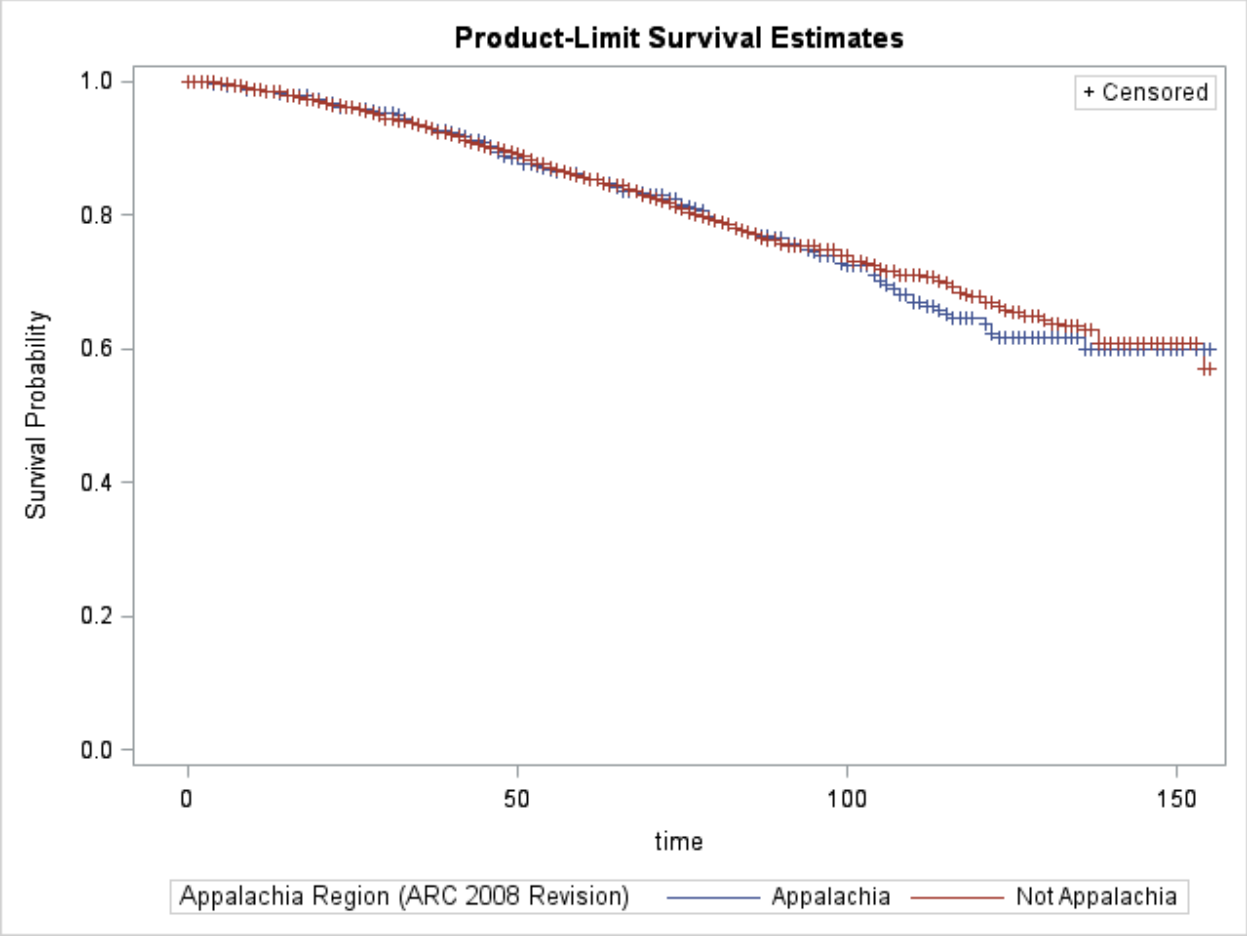
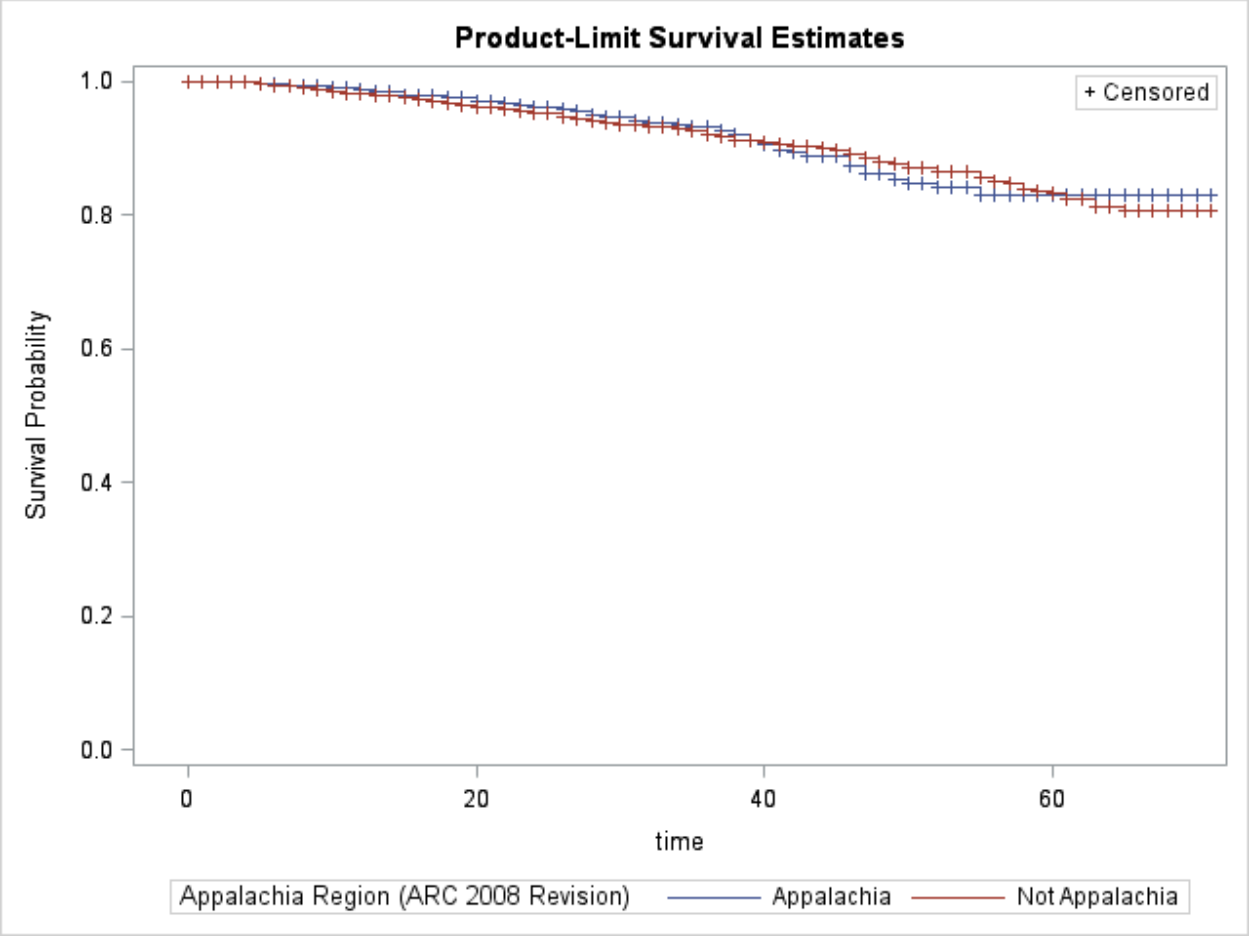


Figure 1.2. The K-M Estimated Survival Curve >2006 (this stratum includes patients who were diagnosed with lung cancer in 2007 through 2013).

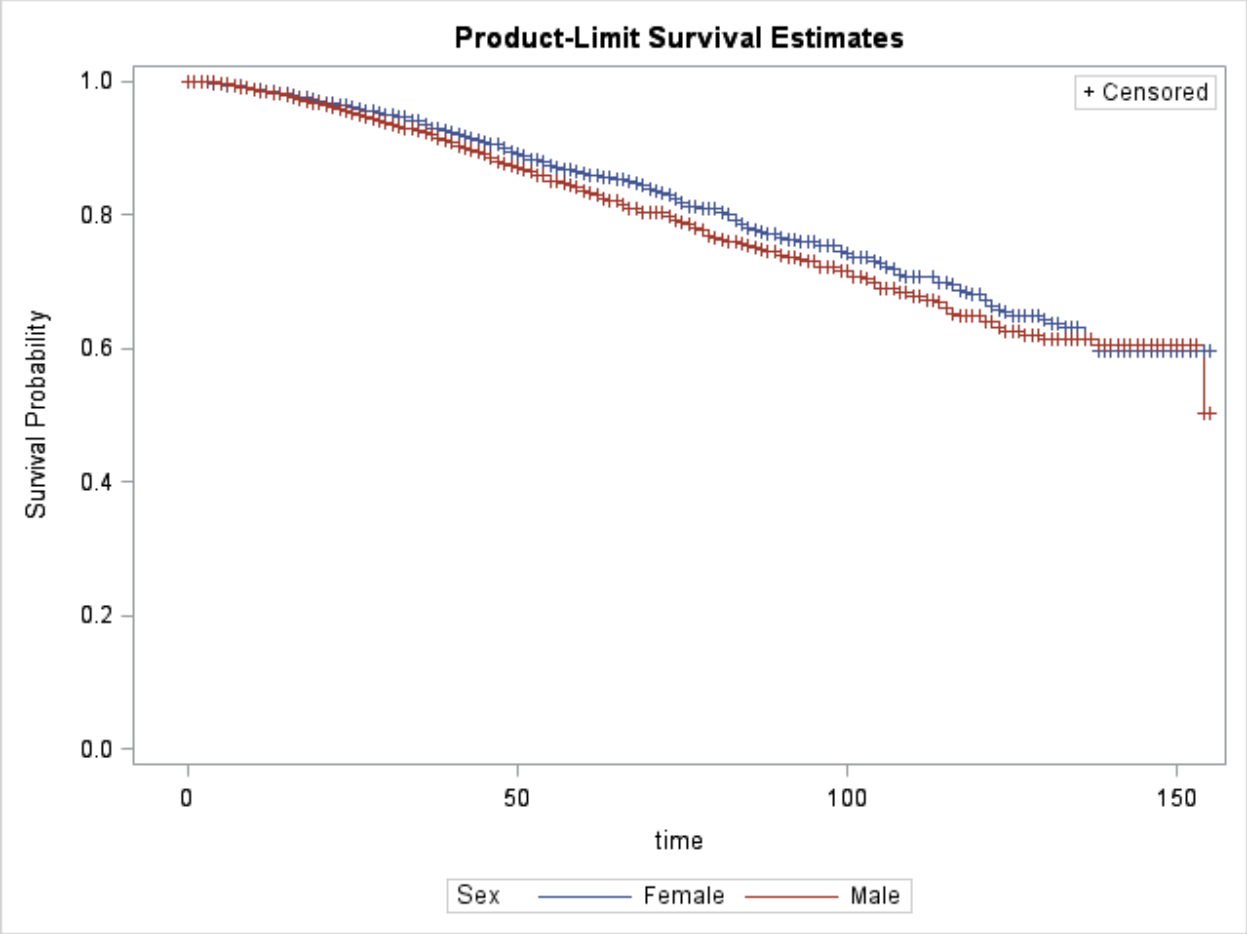
There are no statistically significant differences between two groups, however, we can see that overall survival time is less than in the previous stratum.

Test of Equality over Strata 2007+			
Test	Chi-Square	DF	Pr >
Chi-Square Log-Rank	0.5703	1	0.4502
Wilcoxon	2.2248	1	0.1358
-2Log(LR)	0.8708	1	0.3507



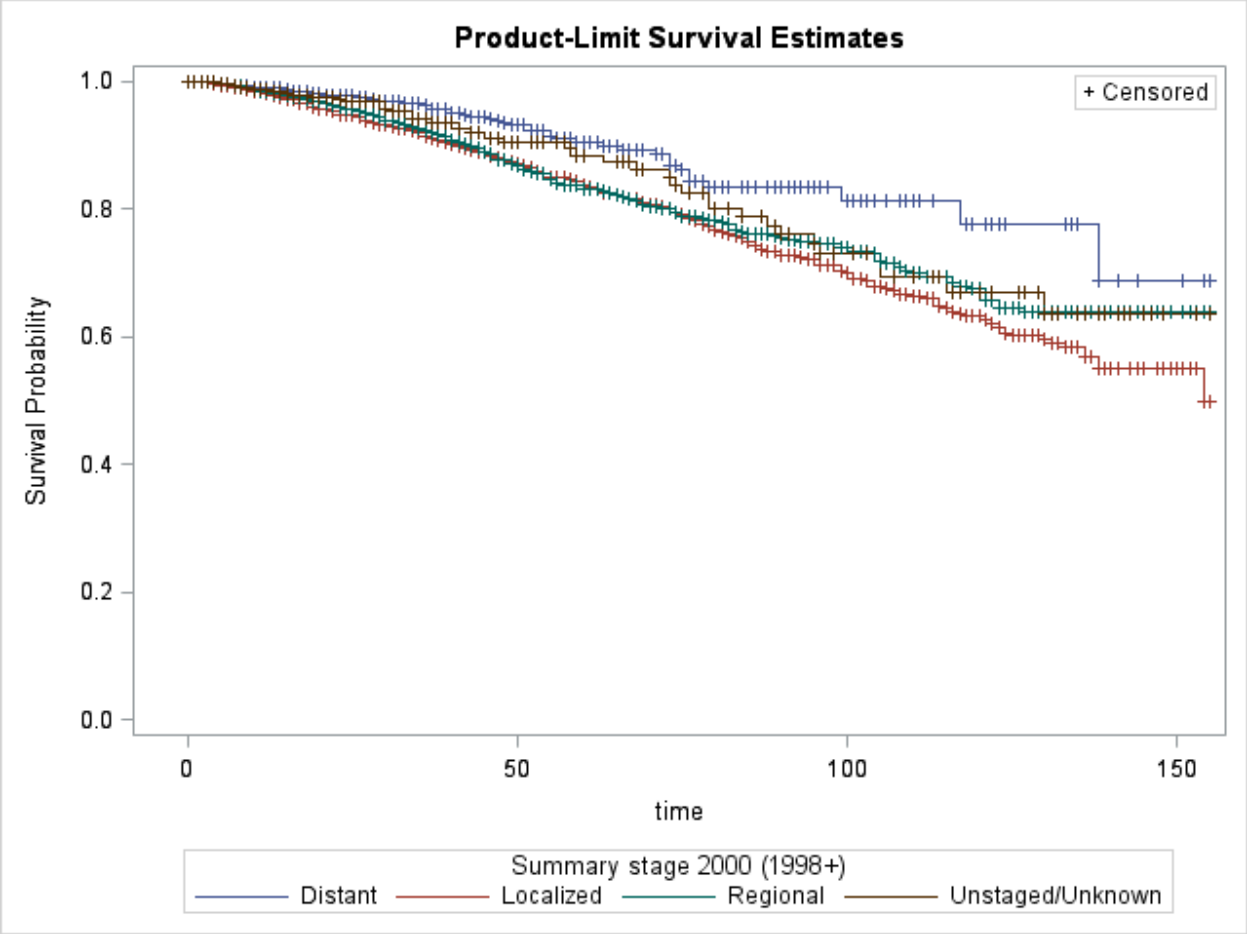
According to the Figure 2, statistically significant differences were observed between females and males with respect to probability of developing subsequent primary cancers over the 155 month time period (59.75% vs. 50.4%, respectively; log-rank test: $\chi 2 = 6.0938$, p=. 0.0136).

Figure 2. The K-M Estimated Survival Curve comparing two gender groups.



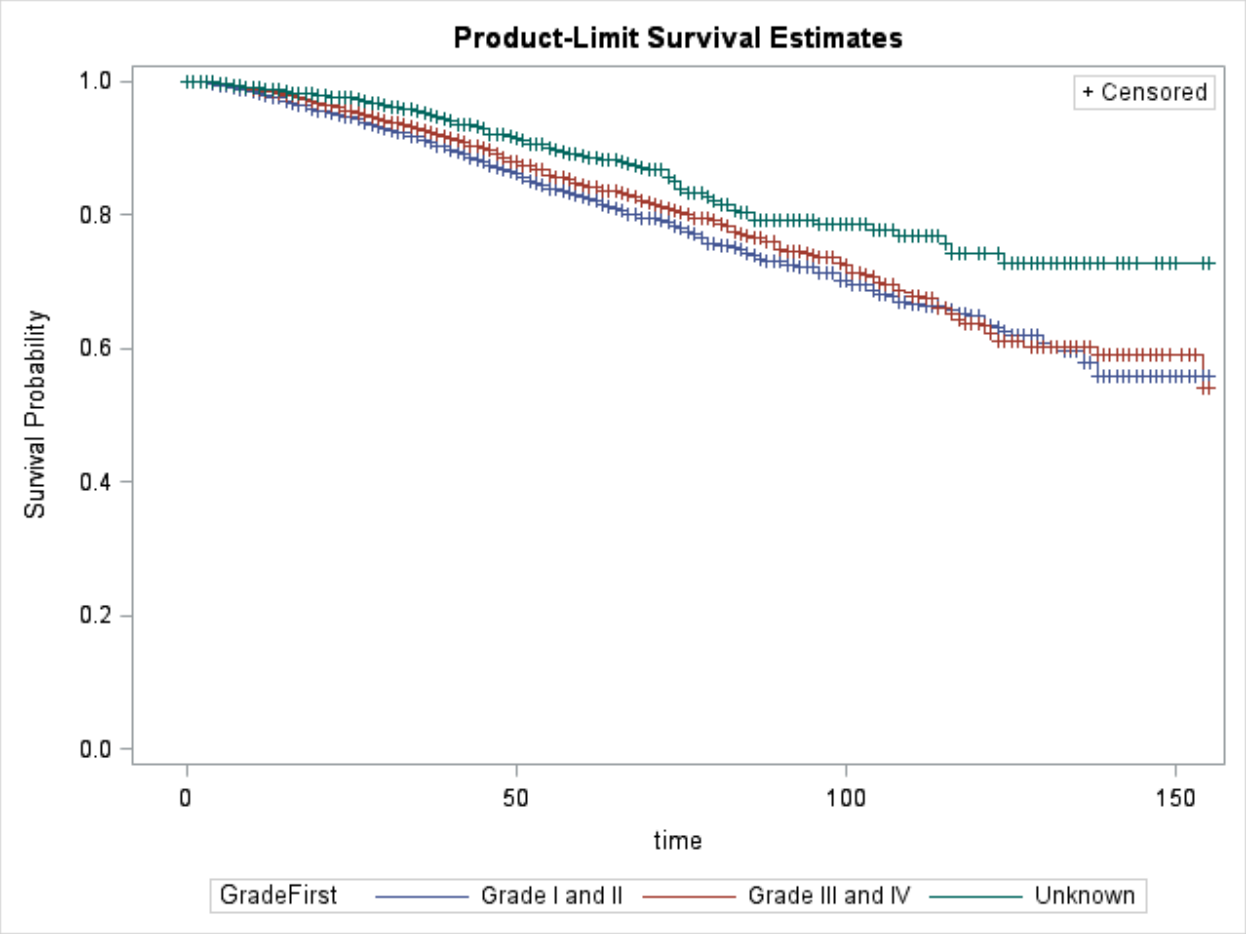
According to Figure 3, statistically significant differences were observed between patients with stage II, III, IV tumors, and those with unknown status, with respect to probability of developing subsequent primary cancers over the 155 month time period (50.01% vs. 63.95% vs. 68.96% vs. 63.77%, respectively; log-rank test: $\chi 2=39.8271$, p=<.0001).

Figure 3. The K-M Estimated Survival Curve comparing four stage groups.



According to Figure 4, statistically significant differences were observed between patients with grade I and II, III and IV tumors, and those with unknown status, with respect to probability of developing subsequent primary cancers over the 155 month time period (55.81% vs. 54.24% vs. 72.78%, respectively; log-rank test: $\chi 2=36.4894$, p=<.0001).

Figure 4. The K-M Estimated Survival Curve comparing three grade groups.



According to Figure 5, statistically significant differences were observed between patients who went through a surgery as the treatment for their initial lung cancer, those who did not have surgery, as well as those with unknown status, with respect to probability of developing subsequent primary cancers over the 155 month time period (54.51% vs. 58.52%, vs. 79.34% respectively; log-rank test: $\chi 2 = 49.3153$, $p = <.0001$).

Figure 5. The K-M Estimated Survival Curve comparing three surgery groups.