2017

DISPARITIES IN STAGE-APPROPRIATE THERAPY FOR RESECTABLE NON-SMALL CELL LUNG CANCER IN KENTUCKY

Jeremiah T. Martin
University of Kentucky, martinjt@somc.org
Author ORCID Identifier:
https://orcid.org/0000-0002-0666-2617
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.251

STUDENT AGREEMENT:

I represent that my thesis or dissertation 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 thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Jeremiah T. Martin, Student

Dr. David Mannino, Major Professor

Dr. David Mannino, Director of Graduate Studies

DISPARITIES IN STAGE-APPROPRIATE THERAPY FOR RESECTABLE NON-SMALL CELL LUNG CANCER IN KENTUCKY

THESIS

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Clinical Research Design (MSCRD)
in the University of Kentucky College of Public Health

By
Jeremiah T. Martin

Lexington, Kentucky

Director: Dr. David Mannino,
Professor & Chair
(Preventive Medicine & Environmental Health)
Lexington, Kentucky

2016

ABSTRACT OF THESIS

DISPARITIES IN STAGE-APPROPRIATE THERAPY FOR RESECTABLE NON-SMALL CELL LUNG CANCER IN KENTUCKY

Lung cancer (NSCLC) is the leading cause of cancer related mortality. Lung cancer screening aims to detect treatable cancers, however survival advantage will only be seen with early and appropriate stage-directed therapy. This study aims to understand recent rates of therapy for early-stage lung cancer in Kentucky, and to explore potential sources of disparities in treatment and outcomes. A Kentucky Cancer Registry query was performed of all NSCLC cases treated in the state from 2005-2014. Of 39,763 lung cancer patients, 10,622 were clinically operable. Of these, overall 40% did not receive surgery, while 16% did not receive any stage-appropriate local therapy. Wide variation was noted in rates of surgery and local therapy at the county level. Increased age, non-private insurance status, non-white race, male gender, and non-married status were less likely to receive surgery. Median survival in patients who underwent surgery was 59.1 months vs 16 months ($p<0.001$). Appropriate stage-directed local therapy is a very important factor in survival of patients with early stage NSCLC. County-level variation in rates of therapy need further study. Demographic factors continue to drive disparities in therapy and outcomes in Kentucky and should inform health policy and ongoing research and education efforts.

KEYWORDS: Lung cancer, Outcomes, Surgery, Cancer Staging, Disparities

Jeremiah T. Martin, MBBCh

(Author)

December 6, 2016

(Date)

DISPARITIES IN STAGE-APPROPRIATE THERAPY
FOR RESECTABLE NON-SMALL CELL LUNG CANCER
IN KENTUCKY

By

Jeremiah Thomas Martin

David Mannino
(Director of Thesis)

David Mannino
(Director of Graduate Studies)

December 6 2016

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge Dr. David Mannino and the School of Public Health at the University of Kentucky for the opportunity to undertake this MSCRD while a faculty member at UK. This was an invaluable opportunity for personal academic development and growth.

I am grateful to Eric Durbin and Bin Huang at the Kentucky Cancer Registry who supported this and other KCR projects with quality data, helpful insight, and patience.

I would like to thank Dr. Victor Ferraris for his mentorship on this project and guidance throughout my early career development as a thoracic surgeon.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

INTRODUCTION

Lung cancer is the leading cause of cancer-related mortality in the United States. It accounts for more cancer-related deaths than the next four causes of cancer combined for both genders.[1] Despite this, the overall survival from lung cancer remains poor and has not advanced significantly over the past several decades.

The primary risk factor for the development of lung cancer is smoking.[2] Genetics, other occupational exposure, and prior cancers also play a role. Lung cancer develops when normal lung tissue becomes abnormal and begins to proliferate. Due to differences in treatment and behavior, lung cancers are divided into “small cell” and “non-small cell lung cancer”. The most common type of lung cancer is non-small cell lung cancer, and is the one which will be discussed in this project.

Lung cancer is generally without symptoms until more advanced stages. The most common stage of presentation of lung cancer is stage 3, at which point treatment is generally supportive with chemotherapy and radiation delivering overall survival of less than 15% at 5 years. Over the past several decades multiple efforts have been made to develop a screening tool for lung cancer, however it was not until the arrival of the low-dose CT screening programs that it was possible to potentially detect early stage lung cancer.[3]

Reduction of mortality from lung cancer is a complex task. It requires risk factor and lifestyle modification for prevention, effective early detection mechanisms to find early-stage cancers, and complex multidisciplinary teams to quickly and effectively treat cancers when found.

When lung cancer is detected or suspected, early referral to a thoracic surgeon is important. The surgeon, in conjunction with the team at a cancer center, will make a determination as to be clinical stage of the patient. In early-stage cancer (1 and 2) the primary treatment should be a local therapy. This is ideally a lobectomy for a surgical candidate, or stereotactic radiation for a nonsurgical patient. Patients with stage 2 cancer may require chemotherapy in the adjuvant setting.

The initial workup of the cancer patient generally includes a brain MRI and PET scan to screen for metastatic disease. In the absence of distant metastasis, examination of the mediastinal lymph nodes checks for central nodal involvement. Finally, if the tumor is anatomically resectable, surgery is recommended if the patient's cardiopulmonary reserve can tolerate this.

This initial decision making pathway is complex. It involves many variables which can be measured, but also relies on the clinical best judgment of the surgical team and staff at the cancer center.

As screening programs become more prevalent, it is likely that more patients with early-stage cancer will be identified. In order for lung cancer screening to therefore demonstrate a survival advantage, it is important that these early-stage patients receive optimal treatment.

The objective of the current study is to

1) **determine the state-wide rates of surgery and local therapy in patients with early-stage lung cancer,**

2) **to investigate the impact of appropriate local therapy on survival,** and

3) **to determine the relative importance of socioeconomic disparities in the receipt of appropriate surgery, and on survival.**

METHODS

Data Source

The Kentucky Cancer Registry (KCR) was chosen as the data source for the study. The KCR is an active surveillance registry of cancer cases in the state of Kentucky. Data are collected from multiple sources. On-site data abstractors collect and complete patient records, pathology information is transmitted electronically to the KCR, and periodic audits and reviews of data completeness and accuracy are performed. KCR records the full address of origin of each patient, and these data were extracted at the individual level: KCR staff collapsed these data to the county level and provided county-level linkage with socioeconomic data from the 2010 census.

University of Kentucky Institutional Review Board Approval was obtained: This study was deemed IRB exempt as a retrospective review of existing data collected in a registry.

Variables and selection criteria

The KCR was queried for records to include years 2005-2014. All patients with Non-Small Cell Lung Cancer diagnosis were extracted (Site Codes c340-c349 define carcinoma of the lung and bronchus; Small cell histology codes 8041-8045 were excluded) Demographic variables included age at diagnosis, year of diagnosis, marital status, gender, race, smoking status, insurance. Socioeconomic variables included county of residency, % of population graduating with high-school diploma in 2010, % of population below the poverty line in 2010. Clinical variables included tumor site and size, clinical T (tumor), clinical N (nodes), clinical M (metastasis), computed best stage based on AJCC clinical and pathological stages, class of case (referral), tumor grade, first course treatment, cancer sequence number, ICD-O-3 histology codes. A composite treatment code was used to determine the initial choice of treatment in each case. Follow-up variables included date of last contact and vital status.

The data were first grouped using a “Clinically Operable” categorical variable (CLINOP). This was constructed using an algorithm that mirrors that used in the cancer clinic by a thoracic surgeon. A patient is presumed to be operable until an exclusion is found. Starting with (M)etastasis, checking (N)odal involvement and finally (T)umor resectability. If a patient was missing data in any or all of these counts, they were given the "benefit of the doubt" which is an accepted standard in the workup of cancer patients. This new CLINOP variable was compared with the KCR's "best stage group", and both were used to select data subsets for further analysis. These are the potentially operable, early stage, cohorts.

The composite treatment variable was now used to stratify the patients into two different comparison groups: The groups that received surgery (SURG) vs no surgery, and the groups that received local therapy (LOCAL) vs no local therapy. Local therapy was defined as surgery or radiation or combination as first-line therapy.

Statistical Methods

All statistical analysis was conducted using the R Environment for Statistical Computing.[4]

Numerical continuous variables were analyzed using descriptive statistics such as median and inter-quartile range. Categorical variables were reported as frequencies and percentages.

After initial data exploration (Figure 1), statistical learning (decision tree) was used to search for potential predictors of whether or not a patient would receive surgery, based on the available input variables. The dataset was divided into a 70/30 training/test dataset and a conditional decision tree was constructed. (Figure 2).

Descriptive statistics were generated for each group (Table 1). Logistic regression was used to measure the associations between demographic and socioeconomic predictors on the receipt of surgery (Table 2). A multivariate logistic model was constructed using forward selection.

Survival was examined using Kaplan-Meier estimates, and stratifying the patients into SURG/LOCAL groups in addition to demonstrating the effects in both the CLINOP group and the KCR Best Stage cohort (Table 3). Cox Proportional Hazards models were constructed to calculate unadjusted univariate Hazard Ratios for long-term survival based on receipt of surgery vs local therapy and a multivariate Cox Proportional Hazards model returned adjusted Hazard Ratios. (Table 4)

Finally, although individual county-level measurements did not reach statistical significance on multivariate modeling, there appear to be wide variations in the rates at which patients receive surgery. These rates were tabulated and graphically represented to facilitate future discussion and study. (Table 5, Figure 3)

RESULTS

Description of the dataset, initial exploration

The KCR query returned 39,763 patients with NSCLC in the study period 2005-2014. The initial stratification revealed cohorts of 10,622 clinically operable patients, and 11,274 KCR best stage. Since the intent was to analyze early within the decision tree, and much of the analysis focused on the clinically operable group.

Within the clinically operable group, overall 40% (4,203) patients did not receive surgery. Stage appropriate local therapy includes radiation to patients who may be poor candidates for surgery however within the clinical stage 1 and 2 group, 16% (1736) did not receive stage appropriate local therapy (Table 1). Although the Registry does not contain details on comorbidities, there remained important differences in demographic and socioeconomic variables between patients who received surgery or local therapy versus suboptimal therapy. Univariate analysis demonstrated that more married patients received surgery or local therapy compared to single or divorced. White race and female gender accounted for more of the surgically and locally treated patients. The presence of private insurance and Medicare accounted for more of the treated patients compared with self-pay, Medicaid, or military or other government insurance. Univariate differences were not identified when stratifying the county of residence as appalacian vs not, rural vs not, or based on percentage of population graduating high school, or percentage of the population below the poverty line.

Figure 1 graphically illustrates some differences between Appalachian Counties versus the remainder of the state. There are expected differences in rates of high school graduation, and percent below the poverty line. Breakdown of insurance coverage is different within these regions, however smoking and surgery rates do not appear visually different.

Statistical Learning

Prior to building a logistic regression model, a conditional decision tree was built including the demographic and socioeconomic inputs to recursively stratify the population based on whether or not surgery was received. Figure 2 illustrates these findings. Age of diagnosis is the most important factor, with a split point of 78 years representing an initial partition of the data between those who did, versus did not receive surgery. Amongst the younger patients, insurance was the next most important partition point for those patients less than 61, while marital status was more important for patients greater than 61. This decision tree was developed on a training dataset derived from a random sample of 80% of the clinically operable patients. The model was tested on the remaining 20% with a 34% overall error rate, and AUC of 0.6745.

Factors associated with Receipt of Surgery

Logistic regression was used to determine the individual and relative strengths of association between the available variables and receipt of surgery. Findings are summarized in Table 2. Univariate odds ratios were generated by running individual regression models for the variable of interest with the outcome of surgery. Forward selection based on reduction of AIC was used to generate a final fitted logistic model. The AUC for the final fitted logistic model was 0.693

Patients who are older were less likely to have received surgery (OR 0.94 per year, 95% CI 0.93-0.94, $p<0.001$). Patients who were married were more likely to have received surgery (OR 1.59 vs single, 95% CI 1.33-1.89, $p<0.001$). Patients with private insurance (OR 1.42) and Medicare (OR 1.42) were more likely than self-pay or uninsured to have received surgery however patients with Medicaid (OR 0.65) or military/government (OR 0.52) insurance were less likely to have received surgery than self-pay or uninsured.

Patients whose cancer diagnosis was rendered at a site other than treating facility were much less likely to have received surgery (OR 0.13, 95% CI 0.07-0.27, $p<0.001$)

There is also a time-related trend in surgery: compared with 2005, the odds of having received surgery in the adjusted model increased to a peak of 2.06 in 2011, before decreasing towards 1.48 in 2014.

Smoking history (OR 0.6, 95% CI 0.5-0.73, $p<0.001$), female gender (OR 1.27, 95% CI 1.15-1.4, $P<0.001$), and white race (OR 1.47, 95% CI 1.22-1.79, 95% CI $P<0.001$) also contributed to the final model.

Geographic socioeconomic factors including Appalachia vs not, urban/rural continuum, % population below poverty line, and % population graduating high school, did not remain significant in the final adjusted model.

Factors associated with Survival

Kaplan-Meier analysis was performed on the derived clinically operable cohort in addition to the KCR “Best Stage” cohort. This analysis was stratified based on whether or not surgery was performed, and whether or not local therapy was used.

In all cases, surgery or local therapy was associated with significant survival differences. In the clinically operable cohort, the median survival in patients who underwent surgery was 59.1 months vs 16 months ($p<0.001$) . The median survival among patients who underwent any local therapy (surgery or radiation) was 44.15 months vs 8.61 months ($p<0.001$). (Table 3)

Cox proportional hazards regression models were performed at the individual variable level and factors found to be significant were chosen and entered into two final adjusted models (Table4). The first model was adjusted for significant covariates, and included the primary outcome of surgery vs not. The hazard ration for surgery of 0.37 (95% CI 0.35-0.39, $p<0.001$) was the strongest association with improved survival when controlling for the other factors in the model. Female gender (HR 0.71, 95% CI 0.67-0.75, $p<0.001$), Insurance Status (Medicaid HR 1.23, 95% CI 1.07-1.42, $p=0.005$), and age at diagnosis (HR 1.02 $p<0.001$) also retained significant independent association with decreased survival in the final adjusted model.

A second multivariable model, adjusted for local therapy vs not, had a hazard ratio for local therapy of 0.32 (95% CI 0.3-0.34, $p<0.001$) which again was the strongest factor associated with survival. Female gender, age at diagnosis, and insurance status again remained significant in the final model. Marital status (Married HR 0.9, 95% CI 0.81-0.99, $p=0.028$) also remained significant in this model.

State-wide variation in rates of surgery

There are differences in the proportions of patients who undergo surgery, local therapy, and survival rates at the county level in the State of Kentucky throughout the study period. In multivariate modeling, these differences were not found to be statistically significant, however this pattern warrants further study.

Findings are visually illustrated in Figure 3. Table 5 identifies counties with the lowest and highest rates of surgery and local therapy and reports overall and surgical survival for these patients. Rates of surgery in early stage lung cancer vary from 36% in Leslie county, to 80% in Robertson county. Meanwhile local therapy ranges from 39% in Elliott county to 100% in Robertson county.

DISCUSSION

Lung cancer is a highly lethal disease. The primary risk factor of smoking is highly prevalent, public awareness still needs work, and many times the disease is not identified until it is significantly advanced. Screening programs using low-dose computed tomography will hopefully affect a stage-shift with earlier detection of lung cancer, however once cancers have been found in order to affect survival we need to ensure that timely and appropriate stage directed therapy is performed.

Researchers are increasingly concerned with disparities in healthcare, and in the case of cancer care disparities a significantly higher mortality is seen in disadvantaged groups. In his editorial, "Deprivation, distance and death and lung cancer," Dr. Peake summarizes recent literature on the topic and makes the point that it must be our, "aim to find ways of ensuring equitable access to the highest quality of care for all patients with lung cancer were ever they live and whether further social background."[5]

This study of registry-level data demonstrates that there are differences in rates of surgery and local therapy throughout the state of Kentucky. In addition, multivariate analysis demonstrates that whether or not surgery or local therapy is received is also affected by gender, race, and insurance disparities. This is in keeping with findings from other studies.[6-17] Interestingly, the current analysis fails to demonstrate significantly different survival in patients from various parts of the state. This may in part be due to the magnitude of the effect of surgery vs. not on survival, which may mask the much smaller effect of geographic disparity. A longer study period with larger patient population size may further help to delineate this. In addition, the current analysis is limited to those patients were diagnosed with clinically stage 1 and 2 cancer and therefore excludes the likelihood that patients in rural centers may present more advanced stages of disease.

The impact of insurance status on outcomes, which include not only receipt of surgery or local therapy but also survival, is an important public health concern. This trend and association which is demonstrated in the analysis of this dataset was also found in analysis of a national sample. Shi and colleagues[7] found in an analysis of the National Cancer Data Base that Medicaid and uninsured patients are at a higher risk mortality than patients with private insurance. They hypothesized that these disparities may be due to different treatments offered based on insurance status which is also seen in this analysis.

Statistical learning techniques demonstrated one of the most important factors in determining whether or not surgery was offered was age, and a split point of 78 years was identified. Recent literature would indicate that older patients can enjoy long term survival after cancer surgery[18], and therefore this should be kept in mind by physicians treating these patients.

The yearly odds of receiving surgery for stage 1 and 2 lung cancer demonstrated improvement until 2011 with a later decline of unclear importance. Factors which likely impact this include changes in insurance policies around that time, and the increasing application of stereotactic radiation (SBRT) for clinical stage 1 and 2 cancers.[19] While the initial outcomes associated with SBRT are promising, the oncologic efficacy remains to be proven in long term followup. This trend should be closely monitored to ensure that patients continue to receive optimal treatment.

The regional variations in cancer surgery rates are likely driven by availability of and access to healthcare resources. Smith and colleagues analyzed the travel patterns of cancer surgery patients in a regionalized system. They found that regionalization had significantly increased the distance that some patients must travel to receive their surgical care.[20] Comprehensive care of the cancer patient requires multidisciplinary evaluation including availability of radiologists, surgeons, medical and radiation oncologists, and a facility that can support all of these providers. In the present area, referral patterns may be driven by insurance, hospital networks, and provider and patient bias. Ideally, networks of high-quality care centers will work together to optimize outcomes.

An analysis of the National Inpatient Sample demonstrated that non-white race and comorbidities contributed to increased likelihood of receiving cancer surgery at low-volume hospitals.[21] This is an important consideration as hospitals that infrequently treat lung cancer may be less likely to have the full support network in place to provide optimal multidisciplinary care. In the state of Kentucky, travel distances involved for patients to reach accredited cancer centers, or academic facilities, may be prohibitive for many with limited means.

Public Health Implications

Disparities in healthcare are a major public health concern. These factors account for up to 25% of the variations in outcome in morbidity and mortality in this country.[22, 23]

The issues stem from many areas and there are individual actionable items. Ongoing efforts in all of these domains will improve outcomes in lung cancer therapy and ultimately survival. These areas, and their particular application to lung cancer in Kentucky, can be broadly broken down as follows:

1. Health System Related
    a. Access
    b. Quality
    c. Regulations, Policies and Systems
2. Patient Related
    a. Patient preferences, compliance
    b. Culture, Lifestyle
    c. Biology and genetics
3. Provider Related
    a. Prejudice, bias
    b. Up-to-date knowledge
    c. Cultural insensitivities

In this analysis of the Kentucky Cancer Registry, there are several statistically significant points that lend to action items and further study.

1) Receipt of Local Therapy or Surgery vs none

   This was the single most important predictor of survival and the rates of local therapy or surgery vary widely based on the county of origin and are affected by insurance status, age, race, gender. In addition to being driven by patient beliefs and education, there is an opportunity to educate providers across the healthcare continuum about the importance of early stage-directed therapy for patients with clinical stage 1 and 2 lung cancer. These efforts might yield the most benefit in counties where the rates of such therapy are low. The precise reasons for county-wide variation in surgery or local therapy cannot be extrapolated from the current data set but deserve further study.

2) Age

   Older patients are less likely to receive curative intent surgery or local therapy. Educating provders and patients that age alone should not exclude from receipt of maximal therapy may decrease the impact of this factor.

3) Insurance

   Paying for care continues to be a major public health problem – a patient's ability to pay will affect their seeking timely care, and their choices when treatment options are presented. Payer status may indirectly affect physician's decision making as cancer care requires a team effort with multiple visits and close followup. Raising the awareness of the impact of this factor on both treatment and survival outcomes is an important first step.

4) Race

   Kentucky's ethnic distribution may make it more difficult for minority groups to seek, understand, and benefit from complex care. Education of providers, and targeting at-risk populations to increase awareness of lung cancer screening programs may help these groups.

Limitations

There are a number of important limitations of the current study. First, this represents a retrospective analysis of registry level data. Much of this data is collected retrospectively from chart review. Pathology information is accurate, however clinical staging information relies on documentation of the physicians caring for the patients, and sometimes this is less accurate. When a patient undergoes surgery, complete pathologic staging is available which is much more accurate but not directly applicable to this study: When the Registry "best stage" variable was used to examine the data, additional stage 1 and 2 patients were found who underwent surgery who were missing clinical staging information. However, to remain as close as possible to what would occur in clinicians office, only available clinical staging information was used.

Survival difference was not noted to be significant when stratifying patients based on county of origin. It is likely that selecting only clinical stage 1 and 2 patients introduced bias in this regard. It is probable that more patients with advanced disease are diagnosed in underserved areas however the purpose of the study was to specifically look at rates of surgery in early-stage patients therefore this was not further explored and warrants further investigation.

Data are not available on some of the important clinical comorbidities which a surgeon may use to determine whether or not to offer surgery. In addition the intrinsic bias of the physician and patient during the clinical encounter cannot be measured or recorded and are absent from this analysis.

Future directions

There are indications from the current study that rates of local therapy are changing over time, and that there are wide variations in rates at the county level. Ongoing surveillance of these datapoints is needed and cancer prevention efforts should continue to focus on at-risk populations including underserved counties.

CONCLUSION

In patients with clinically stage 1 and 2 lung cancer in Kentucky, there is wide variation in the rate of curative-intent locally-directed therapy. Receipt of stage appropriate local therapy is the strongest predictor of survival. Race, Gender, Insurance and Marital Status are important demographic factors that contribute to disparities in treatment and outcome.

TABLES & FIGURES

Table 1: Descriptive statistics, and univariate analysis. Stage appropriate surgery is defined as surgery as part of the initial course of therapy for clinically stage 1 and 2 lung cancer patients. Stage appropriate local therapy includes surgery and radiation therapy as part of the initial course of therapy.

			STAGE-APPROPRIATE SURGERY						STAGE-APPROPRIATE LOCAL THERAPY (RADIATION/SURGERY)					
			CLINOP: Clinically Operable Patients			KCR Best Stage Grouping			CLINOP: Clinically Operable Patients			KCR Best Stage Grouping		
		level	Surgery 0	Surgery 1	p	Surgery 0	Surgery 1	p	Local 0	Local 1	p	Local 0	Local 1	p
n			4203	6419		4011	7263		1736	8886		1733	9541	
Marital Status (%)	Single	1	345 (8.4)	482 (7.7)	<0.001	314 (8.1)	552 (7.8)	<0.001	137 (8.2)	690 (7.9)	<0.001	130 (7.8)	736 (7.9)	<0.001
	Married	2	1913 (46.8)	3740 (59.5)		1821 (46.8)	4239 (59.7)		713 (42.4)	4940 (56.8)		689 (41.2)	5371 (57.6)	
	Separated	3	33 (0.8)	51 (0.8)		31 (0.8)	56 (0.8)		15 (0.9)	69 (0.8)		19 (1.1)	68 (0.7)	
	Divorced	4	616 (15.1)	911 (14.5)		569 (14.6)	1018 (14.3)		264 (15.7)	1263 (14.5)		251 (15.0)	1336 (14.3)	
	Widowed	5	1174 (28.7)	1092 (17.4)		1145 (29.4)	1226 (17.3)		543 (32.3)	1723 (19.8)		575 (34.4)	1796 (19.3)	
	Unmarried	6	10 (0.2)	9 (0.1)		11 (0.3)	8 (0.1)		8 (0.5)	11 (0.1)		9 (0.5)	10 (0.1)	
White Race (%)	Other	0	301 (7.2)	340 (5.3)	<0.001	272 (6.8)	396 (5.5)	0.005	137 (7.9)	504 (5.7)	<0.001	116 (6.7)	552 (5.8)	0.156
	White	1	3902 (92.8)	6079 (94.7)		3739 (93.2)	6867 (94.5)		1599 (92.1)	8382 (94.3)		1617 (93.3)	8989 (94.2)	
Sex (%)	Male	1	2304 (54.8)	3289 (51.2)	<0.001	2182 (54.4)	3781 (52.1)	0.018	945 (54.4)	4648 (52.3)	0.11	931 (53.7)	5032 (52.7)	0.468
	Female	2	1899 (45.2)	3130 (48.8)		1829 (45.6)	3482 (47.9)		791 (45.6)	4238 (47.7)		802 (46.3)	4509 (47.3)	
Age at Diagnosis (median [IQR])			72.00 [65.00, 79.00]	67.00 [60.00, 73.00]	<0.001	73.00 [66.00, 80.00]	67.00 [60.00, 74.00]	<0.001	73.00 [66.00, 81.00]	68.00 [61.00, 75.00]	<0.001	74.00 [66.00, 81.00]	68.00 [61.00, 75.00]	<0.001
Laterality (%)	Right	1	2259 (54.4)	3621 (56.5)	0.037	2167 (54.2)	4108 (56.6)	0.015	936 (55.2)	4944 (55.7)	0.721	928 (53.9)	5347 (56.1)	0.091
	Left	2	1895 (45.6)	2792 (43.5)		1831 (45.8)	3150 (43.4)		759 (44.8)	3928 (44.3)		795 (46.1)	4186 (43.9)	
Diagnosis elsewhere (%)		0	4131 (98.3)	6398 (99.7)	<0.001	3912 (97.5)	7223 (99.4)	<0.001	1689 (97.3)	8840 (99.5)	<0.001	1667 (96.2)	9468 (99.2)	<0.001
		1	72 (1.7)	21 (0.3)		99 (2.5)	40 (0.6)		47 (2.7)	46 (0.5)		66 (3.8)	73 (0.8)	
Insurance (%)	Self-pay, none	0	226 (5.4)	537 (8.4)	<0.001	224 (5.6)	658 (9.1)	<0.001	86 (5.0)	677 (7.6)	<0.001	97 (5.6)	785 (8.2)	<0.001
	Private insurance	1	334 (7.9)	1113 (17.3)		293 (7.3)	1202 (16.5)		114 (6.6)	1333 (15.0)		103 (5.9)	1392 (14.6)	
	Medicaid	2	271 (6.4)	426 (6.6)		237 (5.9)	479 (6.6)		112 (6.5)	585 (6.6)		103 (5.9)	613 (6.4)	
	Medicare	3	3129 (74.4)	4120 (64.2)		3052 (76.1)	4689 (64.6)		1346 (77.5)	5903 (66.4)		1367 (78.9)	6374 (66.8)	
	Military , govt	4	243 (5.8)	223 (3.5)		205 (5.1)	235 (3.2)		78 (4.5)	388 (4.4)		63 (3.6)	377 (4.0)	
Tumor Size, cm (mean (sd))			58.35 (59.57)	33.81 (29.52)	<0.001	46.73 (45.42)	32.59 (26.62)	<0.001	68.58 (67.72)	38.62 (37.91)	<0.001	48.79 (47.30)	35.59 (32.06)	<0.001
Smoking (%)		0	200 (5.7)	438 (8.0)	<0.001	195 (5.9)	414 (6.7)	0.124	109 (7.7)	529 (7.0)	0.384	101 (7.2)	508 (6.3)	0.238
		1	3290 (94.3)	5068 (92.0)		3108 (94.1)	5726 (93.3)		1312 (92.3)	7046 (93.0)		1302 (92.8)	7532 (93.7)	
Year of Diagnosis (%)		2005	245 (5.8)	242 (3.8)	<0.001	306 (7.6)	733 (10.1)	<0.001	115 (6.6)	372 (4.2)	0.001	171 (9.9)	868 (9.1)	0.939
		2006	236 (5.6)	313 (4.9)		306 (7.6)	711 (9.8)		102 (5.9)	447 (5.0)		157 (9.1)	860 (9.0)	
		2007	219 (5.2)	344 (5.4)		327 (8.2)	729 (10.0)		94 (5.4)	469 (5.3)		168 (9.7)	888 (9.3)	
		2008	365 (8.7)	692 (10.8)		365 (9.1)	771 (10.6)		169 (9.7)	888 (10.0)		181 (10.4)	955 (10.0)	
		2009	387 (9.2)	702 (10.9)		376 (9.4)	730 (10.1)		167 (9.6)	922 (10.4)		167 (9.6)	939 (9.8)	
		2010	488 (11.6)	847 (13.2)		415 (10.3)	732 (10.1)		227 (13.1)	1108 (12.5)		183 (10.6)	964 (10.1)	
		2011	507 (12.1)	856 (13.3)		442 (11.0)	751 (10.3)		191 (11.0)	1172 (13.2)		171 (9.9)	1022 (10.7)	
		2012	563 (13.4)	796 (12.4)		475 (11.8)	701 (9.7)		222 (12.8)	1137 (12.8)		174 (10.0)	1002 (10.5)	
		2013	580 (13.8)	815 (12.7)		485 (12.1)	714 (9.8)		219 (12.6)	1176 (13.2)		176 (10.2)	1023 (10.7)	
		2014	613 (14.6)	812 (12.6)		514 (12.8)	691 (9.5)		230 (13.2)	1195 (13.4)		185 (10.7)	1020 (10.7)	
Appalacia (%)		0	2970 (70.7)	4460 (69.5)	0.201	2816 (70.2)	4955 (68.2)	0.031	1188 (68.4)	6242 (70.2)	0.14	1138 (65.7)	6633 (69.5)	0.002
		1	1233 (29.3)	1959 (30.5)		1195 (29.8)	2308 (31.8)		548 (31.6)	2644 (29.8)		595 (34.3)	2908 (30.5)	
Rural/Urban Continuum (%)		1	1366 (32.5)	1863 (29.0)	0.004	1322 (33.0)	2109 (29.0)	<0.001	547 (31.5)	2682 (30.2)	0.077	525 (30.3)	2906 (30.5)	<0.001
		2	470 (11.2)	794 (12.4)		403 (10.0)	925 (12.7)		192 (11.1)	1072 (12.1)		157 (9.1)	1171 (12.3)	
		3	462 (11.0)	737 (11.5)		378 (9.4)	745 (10.3)		203 (11.7)	996 (11.2)		154 (8.9)	969 (10.2)	
		4	150 (3.6)	229 (3.6)		126 (3.1)	282 (3.9)		71 (4.1)	308 (3.5)		60 (3.5)	348 (3.6)	
		5	131 (3.1)	194 (3.0)		107 (2.7)	237 (3.3)		50 (2.9)	275 (3.1)		42 (2.4)	302 (3.2)	
		6	447 (10.6)	793 (12.4)		464 (11.6)	857 (11.8)		165 (9.5)	1075 (12.1)		206 (11.9)	1115 (11.7)	
		7	683 (16.3)	1067 (16.6)		733 (18.3)	1285 (17.7)		300 (17.3)	1450 (16.3)		377 (21.8)	1641 (17.2)	
		8	155 (3.7)	260 (4.1)		158 (3.9)	269 (3.7)		65 (3.7)	350 (3.9)		67 (3.9)	360 (3.8)	
		9	339 (8.1)	482 (7.5)		320 (8.0)	554 (7.6)		143 (8.2)	678 (7.6)		145 (8.4)	729 (7.6)	
Rural (%)		0	2298 (54.7)	3394 (52.9)	0.072	2103 (52.4)	3779 (52.0)	0.699	942 (54.3)	4750 (53.5)	0.555	836 (48.2)	5046 (52.9)	<0.001
		1	1905 (45.3)	3025 (47.1)		1908 (47.6)	3484 (48.0)		794 (45.7)	4136 (46.5)		897 (51.8)	4495 (47.1)	
%pop grad HS (mean (sd))			81.96 (7.41)	81.79 (7.22)	0.246	81.67 (7.47)	81.75 (7.27)	0.571	81.69 (7.50)	81.89 (7.26)	0.286	80.85 (7.66)	81.88 (7.27)	<0.001
% pop below poverty (mean (sd))			18.96 (5.72)	19.01 (5.73)	0.656	19.10 (5.86)	19.08 (5.81)	0.895	19.18 (5.80)	18.95 (5.71)	0.125	19.65 (6.06)	18.99 (5.78)	<0.001

Table 2:

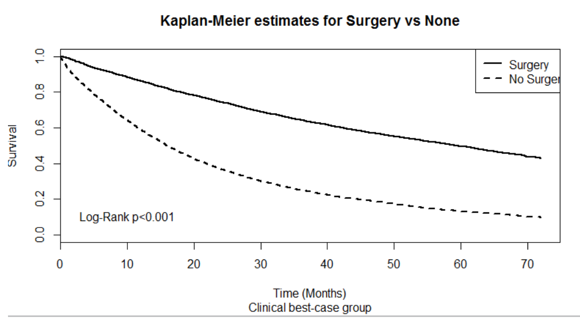
Unadjusted and multivariate adjusted odds ratios for receipt of surgery. The AUC for the fitted model is 0.693

		Unadjusted			Adjusted		
		OR	95% CI	p-value	OR	95% CI	p-value
	Age	0.94	(0.94 - 0.95)	<0.001	0.94	(0.93 - 0.94)	<0.001
Marital Status (%)	*Single*	*ref*			*ref*		
	Married	1.37	(1.17 - 1.61)	<0.001	1.59	(1.33 - 1.89)	<0.001
	Separated	1.18	(0.71 - 1.95)	0.526	1.03	(0.6 - 1.76)	0.914
	Divorced	1.01	(0.84 - 1.22)	0.881	1	(0.82 - 1.23)	0.982
	Widowed	0.64	(0.54 - 0.76)	<0.001	0.96	(0.79 - 1.17)	0.660
	Unmarried	0.53	(0.21 - 1.36)	0.189	0.53	(0.2 - 1.42)	0.205
Insurance (%)	*Self-pay, none*	*ref*			*ref*		
	Private insurance	1.47	(1.19 - 1.83)	<0.001	1.42	(1.13 - 1.79)	0.003
	Medicaid	0.63	(0.5 - 0.8)	<0.001	0.65	(0.51 - 0.84)	0.001
	Medicare	0.55	(0.46 - 0.66)	<0.001	1.22	(1 - 1.49)	0.054
	Military , govt	0.31	(0.22 - 0.45)	<0.001	0.52	(0.35 - 0.75)	0.001
Diagnosis Elsewhere		0.14	(0.07 - 0.27)	<0.001	0.13	(0.07 - 0.27)	<0.001
Year of Diagnosis	*2005*	*ref*			*ref*		
	2006	1.23	(0.94 - 1.62)	0.130	1.27	(0.95 - 1.7)	0.101
	2007	1.54	(1.17 - 2.02)	0.002	1.58	(1.18 - 2.1)	0.002
	2008	1.81	(1.42 - 2.3)	<0.001	1.88	(1.46 - 2.42)	<0.001
	2009	1.7	(1.34 - 2.16)	<0.001	1.88	(1.46 - 2.42)	<0.001
	2010	1.82	(1.45 - 2.3)	<0.001	1.92	(1.5 - 2.45)	<0.001
	2011	1.83	(1.45 - 2.3)	<0.001	2.06	(1.62 - 2.63)	<0.001
	2012	1.44	(1.14 - 1.81)	0.002	1.47	(1.16 - 1.87)	0.002
	2013	1.49	(1.19 - 1.87)	0.001	1.59	(1.25 - 2.03)	<0.001
	2014	1.36	(1.08 - 1.7)	0.008	1.48	(1.17 - 1.88)	0.001
Smoking History		0.68	(0.57 - 0.82)	<0.001	0.6	(0.5 - 0.73)	<0.001
Female Gender		1.15	(1.05 - 1.25)	0.002	1.27	(1.15 - 1.4)	<0.001
White Race		1.47	(1.23 - 1.76)	<0.001	1.47	(1.22 - 1.79)	<0.001
Left sided tumor		0.9	(0.83 - 0.98)	0.020	0.91	(0.83 - 1)	0.044
% pop below poverty		1	(0.99 - 1.01)	0.537	0.99	(0.99 - 1)	0.123
% pop grad HS		0.99	(0.99 - 1)	0.079			
Urban / Rural Continuum	*1*	*ref*					
	2	1.23	(1.06 - 1.43)	0.006			
	3	1.13	(0.97 - 1.31)	0.105			
	4	1.1	(0.86 - 1.39)	0.450			
	5	0.99	(0.77 - 1.27)	0.961			
	6	1.38	(1.19 - 1.61)	<0.001			
	7	1.16	(1.02 - 1.33)	0.025			
	8	1.31	(1.04 - 1.67)	0.025			
	9	1.04	(0.88 - 1.24)	0.643			
Appalacia vs not		1.09	(0.99 - 1.2)	0.073			

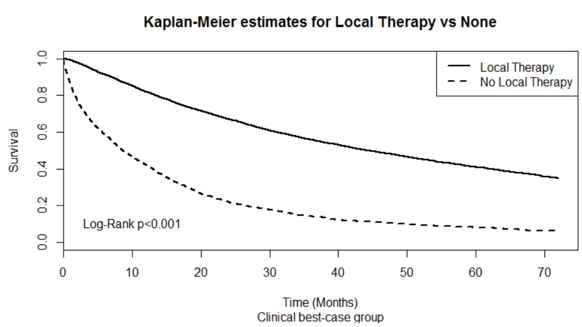
Table 3:

Kaplan-Meier survival analysis. Curves and median survivals are shown for surgery vs not, and local therapy vs not in both the clinically operable group (CLINOP) and the KCR “Best Stage” group.

Clinically Operable Cohort

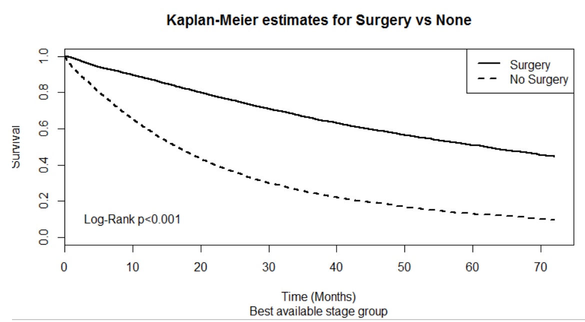


	n	events	median
surg=0	4203	3344	16
surg=1	6419	3111	59.1

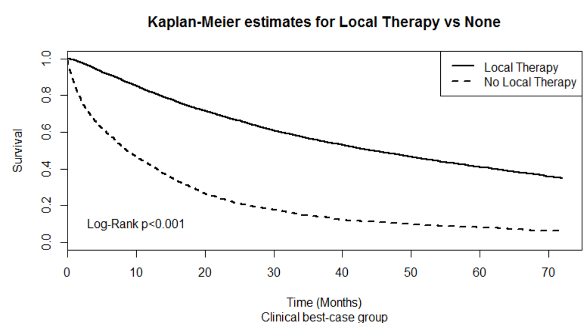


	n	events	median
local=0	1736	1531	8.61
local=1	8886	4924	44.15

KCR Best Stage Cohort



	n	events	median
surg=0	4011	3253	16.4
surg=1	7263	3760	61.8



	n	events	median
local=0	1733	1551	9.8
local=1	9541	5462	48.2

Table 4:

Unadjusted and adjusted multivariate cox proportional hazard models for survival.

		Unadjusted			Adjusted, inc. surgery			Adjusted, inc. local therapy		
		HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
Surgery vs No Surgery		0.33	(0.31 - 0.34)	<0.001	0.37	(0.35 - 0.39)	<0.001			
Local Therapy vs None		0.28	(0.26 - 0.29)	<0.001				0.32	(0.3 - 0.34)	<0.001
Urban / Rural Continuum	*1*	*ref*								
	2	1.02	(0.94 - 1.11)	0.582						
	3	0.96	(0.88 - 1.04)	0.318						
	4	1.16	(1.02 - 1.33)	0.029						
	5	1.06	(0.91 - 1.22)	0.454						
	6	1.02	(0.94 - 1.11)	0.595						
	7	1.06	(0.98 - 1.14)	0.132						
	8	1.00	(0.88 - 1.14)	0.995						
	9	1.12	(1.01 - 1.23)	0.028						
Appalacia vs Not		1.09	(1.03 - 1.15)	0.002	1.07	(0.99 - 1.17)	0.092	1.06	(0.98 - 1.15)	0.139
Female vs Male		0.74	(0.7 - 0.78)	<0.001	0.71	(0.67 - 0.75)	<0.001	0.7	(0.67 - 0.74)	<0.001
Age at Diagnosis		1.03	(1.03 - 1.04)	<0.001	1.02	(1.02 - 1.02)	<0.001	1.03	(1.02 - 1.03)	<0.001
% pop grad HS		1.00	(0.99 - 1)	0.003	1.00	(0.99 - 1)	0.182	1.00	(0.99 - 1)	0.319
% pop below poverty		1.01	(1 - 1.01)	0.02	1.00	(0.99 - 1.01)	0.475	1.00	(0.99 - 1.01)	0.567
White Race vs NonWhite		1.04	(0.94 - 1.16)	0.428	1.04	(0.93 - 1.16)	0.485	1.04	(0.94 - 1.17)	0.431
Diagnosis elsewhere vs not		1.63	(1.28 - 2.09)	<0.001	1.24	(0.96 - 1.61)	0.1	1.14	(0.88 - 1.48)	0.31
Insurance	*Self-pay, none*	*ref*			*ref*			*ref*		
	Private insurance	0.81	(0.72 - 0.92)	0.001	0.86	(0.75 - 0.97)	0.017	0.83	(0.73 - 0.94)	0.004
	Medicaid	1.34	(1.17 - 1.54)	<0.001	1.23	(1.07 - 1.42)	0.005	1.26	(1.09 - 1.45)	0.001
	Medicare	1.55	(1.4 - 1.72)	<0.001	1.06	(0.95 - 1.18)	0.333	1.01	(0.91 - 1.13)	0.842
	Military , govt	1.55	(1.33 - 1.8)	<0.001	0.95	(0.82 - 1.12)	0.561	1.05	(0.9 - 1.23)	0.506
Marital Status	*Single*	*ref*			*ref*			*ref*		
	Married	0.95	(0.86 - 1.05)	0.299	0.94	(0.85 - 1.03)	0.196	0.9	(0.81 - 0.99)	0.028
	Separated	1.04	(0.77 - 1.4)	0.816	1.13	(0.84 - 1.52)	0.432	1.06	(0.79 - 1.43)	0.696
	Divorced	1.10	(0.98 - 1.23)	0.093	1.14	(1.02 - 1.27)	0.024	1.10	(0.99 - 1.23)	0.086
	Widowed	1.31	(1.19 - 1.46)	<0.001	1.11	(0.99 - 1.24)	0.065	1.08	(0.97 - 1.2)	0.173
	Unmarried	1.34	(0.72 - 2.51)	0.36	1.07	(0.57 - 2.01)	0.827	1.00	(0.53 - 1.87)	0.992

Table 5: Variation in rates of surgery and local therapy by county. Overall survival and survival in patients treated with surgery is reported as months (median).

						SURGERY	LOCAL THERAPY	Overall Survival			Surgical Survival		
						rate	rate	n	events	median	n	events	median
Lowest surgery rates	21131	Leslie	Kentucky River	Rural	9 Appalachia	0.364	0.697	33	22	24.4	12	5	NA
	21063	Elliott	Fivco	Rural	9 Appalachia	0.391	0.609	23	19	24.4	9	5	47.5
	21105	Hickman	Purchase	Rural	9 Non-Appalachia	0.429	0.786	14	13	20.5	6	6	30
	21115	Johnson	Big Sandy	Rural	7 Appalachia	0.438	0.672	64	42	29.5	28	13	47.9
	21039	Carlisle	Purchase	Rural	9 Non-Appalachia	0.455	0.818	11	9	13.8	5	3	68.4
Highest Surgery Rates	21213	Simpson	Barren River	Rural	6 Non-Appalachia	0.745	0.836	55	36	34.2	41	22	53.7
	21009	Barren	Barren River	Rural	6 Non-Appalachia	0.750	0.940	100	54	48.7	75	34	60.2
	21031	Butler	Barren River	Rural	8 Non-Appalachia	0.750	0.861	36	22	30.7	27	13	57.7
	21077	Gallatin	Northern Kentucky	Urban	1 Non-Appalachia	0.783	0.870	23	13	50.4	18	8	79.1
	21201	Robertson	Buffalo Trace	Rural	8 Appalachia	0.800	1.000	5	3	46.8	4	2	54.5
Lowest Local Therapy Rates	21063	Elliott	Fivco	Rural	9 Appalachia	0.391	0.609	23	19	24.4	9	5	47.5
	21075	Fulton	Purchase	Rural	7 Non-Appalachia	0.667	0.667	18	11	36.8	12	6	63.4
	21115	Johnson	Big Sandy	Rural	7 Appalachia	0.438	0.672	64	42	29.5	28	13	47.9
	21159	Martin	Big Sandy	Rural	8 Appalachia	0.500	0.692	26	15	32.8	13	4	NA
	21087	Green	Lake Cumberland	Rural	8 Appalachia	0.478	0.696	23	14	49.3	11	4	81.2
Highest Local Therapy Rates	21033	Caldwell	Pennyrile	Rural	6 Non-Appalachia	0.571	0.971	35	26	22.4	20	15	28.9
	21143	Lyon	Pennyrile	Rural	8 Non-Appalachia	0.667	0.972	36	23	30.7	24	14	47.1
	21041	Carroll	Northern Kentucky	Rural	6 Non-Appalachia	0.649	0.973	37	21	34.5	24	10	62.5
	21055	Crittenden	Pennyrile	Rural	6 Non-Appalachia	0.647	1.000	17	11	33.6	11	6	33.6
	21201	Robertson	Buffalo Trace	Rural	8 Appalachia	0.800	1.000	5	3	46.8	4	2	54.5

Figure 1

Graphical exploration of some of the factors that may be important in determining outcomes differences between patients who live in appalacian counties versus not.

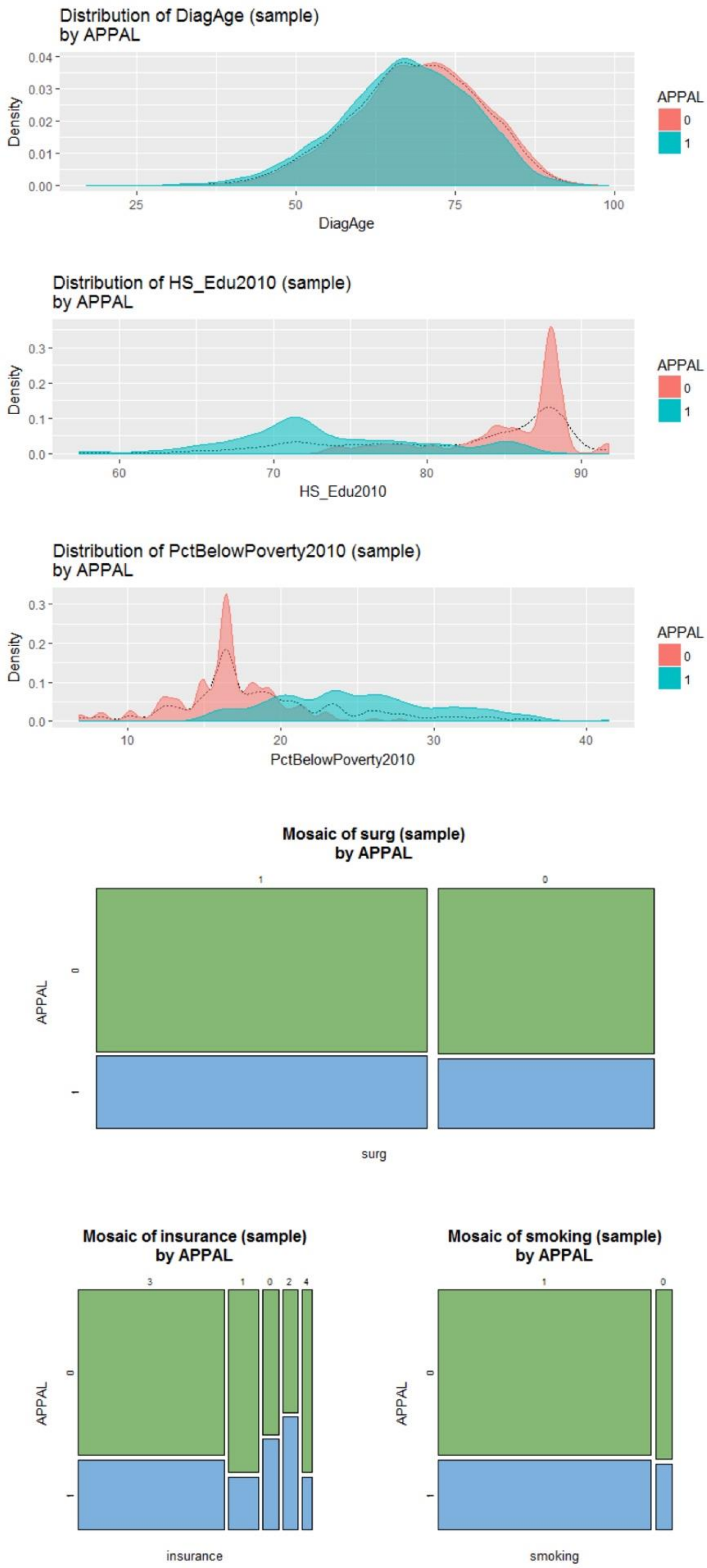


Figure 2: Conditional decision tree: Developed on 80% of the clinically operable dataset, this statistical learning algorithm illustrates variables of importance, and potential split-points, in determining receipt of surgery vs not. When tested on the remaining 20% dataset, the predictive accuracy was 76% with an AUC of 0.6745

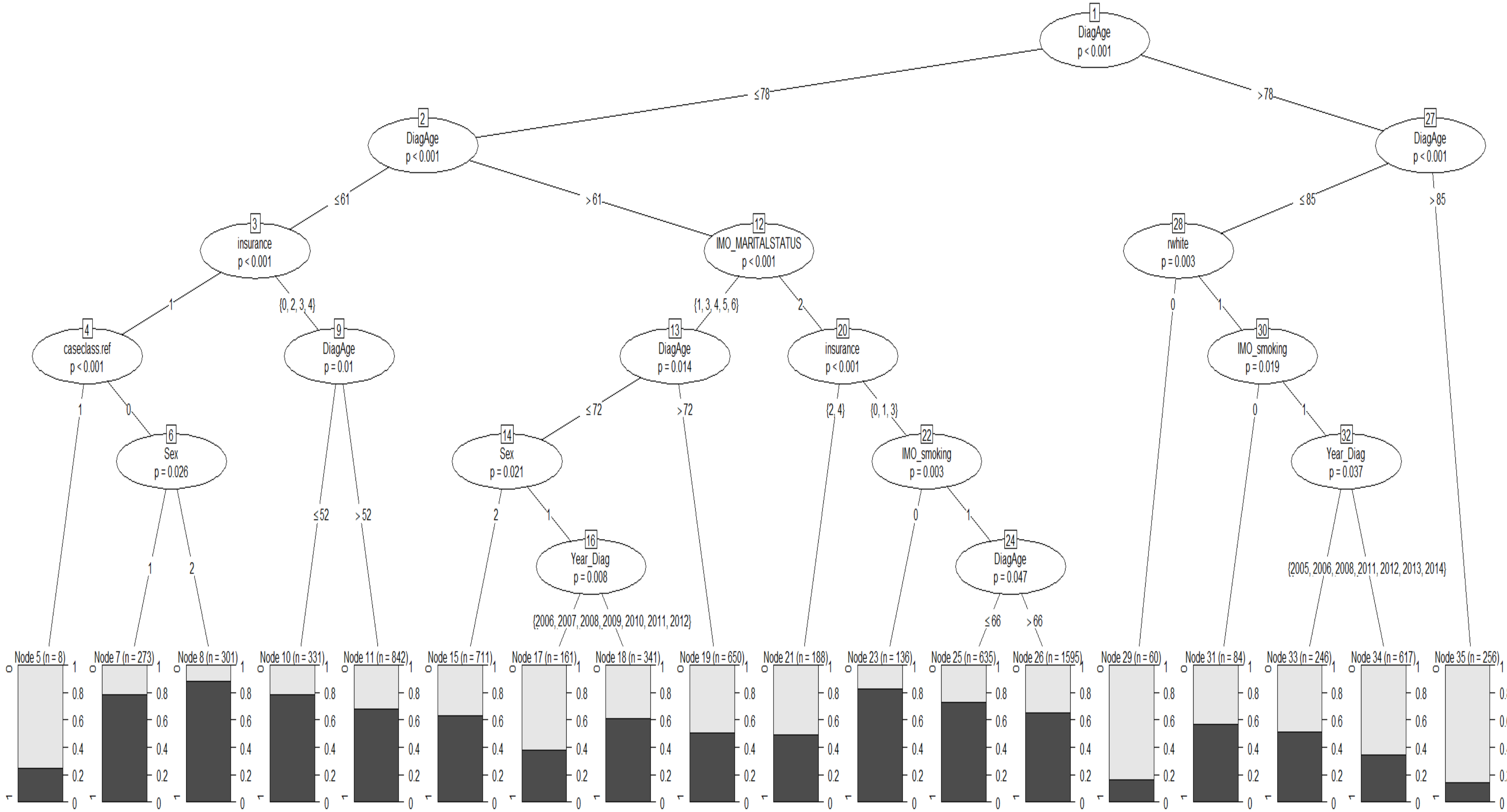
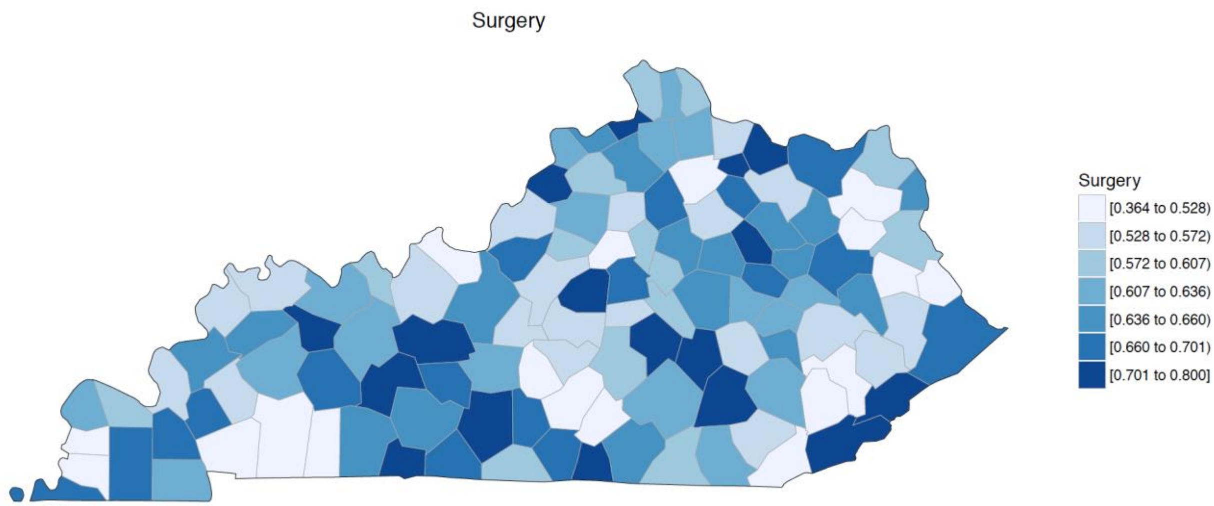
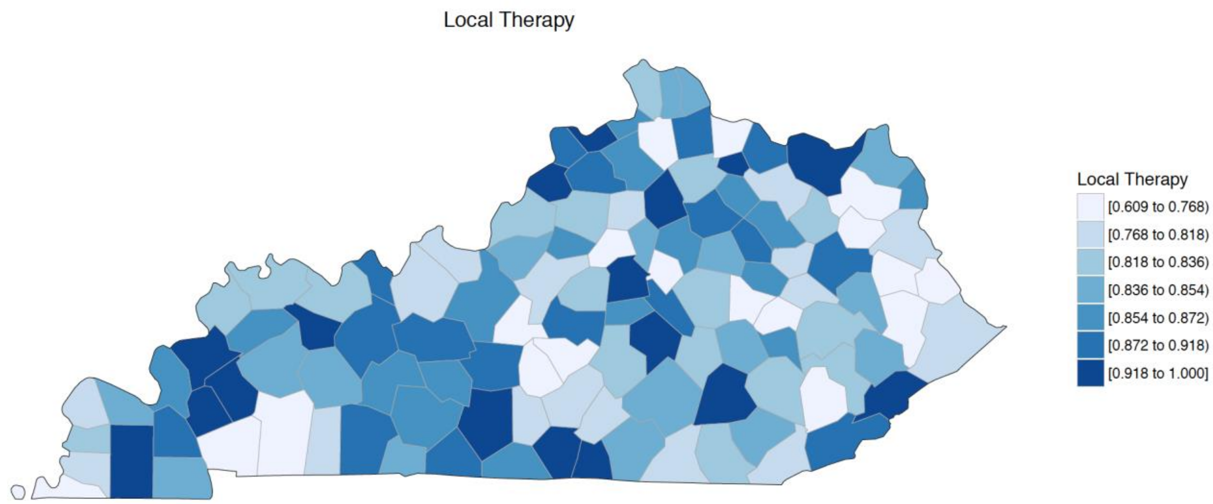


Figure 3

Choropleth representation of rates of surgery and local therapy by county in the state of Kentucky over the 10 year study period.





REFERENCES

1. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2016.* CA Cancer J Clin, 2016. **66**(1): p. 7-30.
2. Boyle, J.M., et al., *Smoking history predicts for increased risk of second primary lung cancer: a comprehensive analysis.* Cancer, 2015. **121**(4): p. 598-604.
3. National Lung Screening Trial Research, T., et al., *Reduced lung-cancer mortality with low-dose computed tomographic screening.* N Engl J Med, 2011. **365**(5): p. 395-409.
4. R Development Core Team, *R: A language and environment for statistical computing*. 2016, R Foundation for Statistical Computing: Vienna, Austria.
5. Peake, M.D., *Deprivation, distance and death in lung cancer.* Thorax, 2015. **70**(2): p. 108-9.
6. Koshy, M., et al., *Disparities in treatment of patients with inoperable stage I non-small cell lung cancer: a population-based analysis.* J Thorac Oncol, 2015. **10**(2): p. 264-71.
7. Shi, R., et al., *The Effect of Payer Status on Survival of Patients with Stage I/II Non-small Cell Lung Cancer: NCDB 1998-2011.* Anticancer Res, 2016. **36**(1): p. 319-26.
8. Okamoto, J., H. Kubokura, and J. Usuda, *Factors Determining the Choice of Surgical Procedure in Elderly Patients with Non-Small Cell Lung Cancer.* Ann Thorac Cardiovasc Surg, 2016. **22**(3): p. 131-8.
9. Williams, C.D., et al., *Influence of comorbidity on racial differences in receipt of surgery among US veterans with early-stage non-small-cell lung cancer.* J Clin Oncol, 2013. **31**(4): p. 475-81.
10. Samson, P., et al., *The National Surgical Quality Improvement Program risk calculator does not adequately stratify risk for patients with clinical stage I non-small cell lung cancer.* J Thorac Cardiovasc Surg, 2016. **151**(3): p. 697-705 e1.
11. Melvan, J.N., et al., *Nonclinical Factors Associated with 30-Day Mortality after Lung Cancer Resection: An Analysis of 215,000 Patients Using the National Cancer Data Base.* J Am Coll Surg, 2015. **221**(2): p. 550-63.
12. Sher, D.J., M.J. Liptay, and M.J. Fidler, *Prevalence and predictors of neoadjuvant therapy for stage IIIA non-small cell lung cancer in the National Cancer Database: importance of socioeconomic status and treating institution.* Int J Radiat Oncol Biol Phys, 2014. **89**(2): p. 303-12.

13. Weksler, B., et al., *Racial and Ethnic Differences in Lung Cancer Surgical Stage: An STS Database Study.* Thorac Cardiovasc Surg, 2015. **63**(7): p. 538-43.
14. Poghosyan, H., et al., *Racial Disparities in Health-Related Quality of Life After Lung Cancer Surgery: Findings From the Cancer Care Outcomes Research and Surveillance Consortium.* J Thorac Oncol, 2015. **10**(10): p. 1404-12.
15. Corso, C.D., et al., *Racial disparities in the use of SBRT for treating early-stage lung cancer.* Lung Cancer, 2015. **89**(2): p. 133-8.
16. Harrison, M.A., et al., *Racial disparity in in-hospital mortality after lobectomy for lung cancer.* Am J Surg, 2015. **209**(4): p. 652-8.
17. Tong, B.C., et al., *Sex differences in early outcomes after lung cancer resection: analysis of the Society of Thoracic Surgeons General Thoracic Database.* J Thorac Cardiovasc Surg, 2014. **148**(1): p. 13-8.
18. Eguchi, T., et al., *Impact of Increasing Age on Cause-Specific Mortality and Morbidity in Patients With Stage I Non–Small-Cell Lung Cancer: A Competing Risks Analysis.* Journal of Clinical Oncology. **0**(0): p. JCO690834.
19. Crabtree, T.D., et al., *Analysis of first recurrence and survival in patients with stage I non-small cell lung cancer treated with surgical resection or stereotactic radiation therapy.* J Thorac Cardiovasc Surg, 2014. **147**(4): p. 1183-1191; discussion 1191-2.
20. Smith, A.K., et al., *Travel patterns of cancer surgery patients in a regionalized system.* J Surg Res, 2015. **199**(1): p. 97-105.
21. Al-Refaie, W.B., et al., *Who receives their complex cancer surgery at low-volume hospitals?* J Am Coll Surg, 2012. **214**(1): p. 81-7.
22. McGinnis, J.M. and W.H. Foege, *Actual causes of death in the United States.* JAMA, 1993. **270**(18): p. 2207-12.
23. Satcher, D. and E.J. Higginbotham, *The public health approach to eliminating disparities in health.* Am J Public Health, 2008. **98**(3): p. 400-3.

APPENDIX

1. Per-County Rates of surgery, local therapy, and survival
2. R-Script for Data Analysis
3. IRB document

						SURGERY	LOCAL THERAPY	Overall Survival			Overall Surgical Survival		
						rate	rate	n	events	median	n	events	median
21001	Adair	Lake Cumberland	Rural	7	Appalachia	0.465	0.791	43	28	30.3	20	9	96.9
21003	Allen	Barren River	Rural	6	Non-Appalachia	0.661	0.911	56	42	26.7	37	25	36.5
21005	Anderson	Bluegrass	Rural	6	Non-Appalachia	0.479	0.750	48	27	46	23	7	94
21007	Ballard	Purchase	Rural	9	Non-Appalachia	0.625	0.813	16	9	35	10	5	35
21009	Barren	Barren River	Rural	6	Non-Appalachia	0.750	0.940	100	54	48.7	75	34	60.2
21011	Bath	Gateway	Rural	8	Appalachia	0.657	0.857	35	19	41.6	23	8	NA
21013	Bell	Cumberland Valley	Rural	7	Appalachia	0.470	0.803	66	39	40.8	31	11	NA
21015	Boone	Northern Kentucky	Urban	1	Non-Appalachia	0.572	0.827	173	101	34.9	99	46	58.6
21017	Bourbon	Bluegrass	Urban	2	Non-Appalachia	0.571	0.878	49	27	40.6	28	9	NA
21019	Boyd	Fivco	Urban	2	Appalachia	0.648	0.855	179	114	42.4	116	64	54.7
21021	Boyle	Bluegrass	Rural	7	Non-Appalachia	0.563	0.854	48	24	34.2	27	11	NA
21023	Bracken	Buffalo Trace	Urban	1	Non-Appalachia	0.538	0.731	26	16	25.2	14	6	54
21025	Breathitt	Kentucky River	Rural	7	Appalachia	0.557	0.836	61	43	26.8	34	22	45.5
21027	Breckinridge	Lincoln Trail	Rural	8	Non-Appalachia	0.561	0.772	57	34	43.3	32	13	NA
21029	Bullitt	Kipda	Urban	1	Non-Appalachia	0.663	0.853	190	101	43.5	126	57	65.3
21031	Butler	Barren River	Rural	8	Non-Appalachia	0.750	0.861	36	22	30.7	27	13	57.7
21033	Caldwell	Pennyrile	Rural	6	Non-Appalachia	0.571	0.971	35	26	22.4	20	15	28.9
21035	Calloway	Purchase	Rural	7	Non-Appalachia	0.607	0.845	84	55	28.7	51	26	61.7
21037	Campbell	Northern Kentucky	Urban	1	Non-Appalachia	0.582	0.851	201	138	28.8	117	69	45.4
21039	Carlisle	Purchase	Rural	9	Non-Appalachia	0.455	0.818	11	9	13.8	5	3	68.4
21041	Carroll	Northern Kentucky	Rural	6	Non-Appalachia	0.649	0.973	37	21	34.5	24	10	62.5
21043	Carter	Fivco	Rural	6	Appalachia	0.524	0.744	82	60	25.2	43	23	50
21045	Casey	Lake Cumberland	Rural	9	Appalachia	0.583	0.833	36	22	33	21	9	42.5
21047	Christian	Pennyrile	Urban	3	Non-Appalachia	0.509	0.716	116	81	30	59	33	52.3
21049	Clark	Bluegrass	Urban	2	Appalachia	0.645	0.849	93	60	28.2	60	31	45.1
21051	Clay	Cumberland Valley	Rural	7	Appalachia	0.611	0.833	36	21	34.5	22	8	NA
21053	Clinton	Lake Cumberland	Rural	9	Appalachia	0.714	0.964	28	16	33.5	20	11	33.6
21055	Crittenden	Pennyrile	Rural	6	Non-Appalachia	0.647	1.000	17	11	33.6	11	6	33.6
21057	Cumberland	Lake Cumberland	Rural	9	Appalachia	0.679	0.929	28	17	42.1	19	9	55.2
21059	Daviess	Green River	Urban	3	Non-Appalachia	0.610	0.830	364	235	37	222	116	61.7
21061	Edmonson	Barren River	Urban	3	Appalachia	0.683	0.854	41	24	36.5	28	14	39.9
21063	Elliott	Fivco	Rural	9	Appalachia	0.391	0.609	23	19	24.4	9	5	47.5
21065	Estill	Bluegrass	Rural	6	Appalachia	0.623	0.717	53	30	35.2	33	13	68.8
21067	Fayette	Bluegrass	Urban	2	Non-Appalachia	0.653	0.855	475	269	36.6	310	141	54
21069	Fleming	Buffalo Trace	Rural	7	Appalachia	0.571	0.800	35	19	33.5	20	7	NA
21071	Floyd	Big Sandy	Rural	7	Appalachia	0.540	0.758	161	104	28.5	87	45	56.6
21073	Franklin	Bluegrass	Rural	4	Non-Appalachia	0.550	0.775	120	83	23.1	66	40	34.7
21075	Fulton	Purchase	Rural	7	Non-Appalachia	0.667	0.667	18	11	36.8	12	6	63.4
21077	Gallatin	Northern Kentucky	Urban	1	Non-Appalachia	0.783	0.870	23	13	50.4	18	8	79.1
21079	Garrard	Bluegrass	Rural	6	Appalachia	0.591	0.909	44	27	47.2	26	15	57.4
21081	Grant	Northern Kentucky	Urban	1	Non-Appalachia	0.620	0.740	50	35	25.2	31	18	53.9
21083	Graves	Purchase	Rural	7	Non-Appalachia	0.685	0.959	73	48	43.8	50	31	53.9
21085	Grayson	Lincoln Trail	Rural	6	Non-Appalachia	0.701	0.883	77	45	41.3	54	28	57.2
21087	Green	Lake Cumberland	Rural	8	Appalachia	0.478	0.696	23	14	49.3	11	4	81.2
21089	Greenup	Fivco	Urban	2	Appalachia	0.588	0.840	119	81	30.5	70	40	44.4
21091	Hancock	Green River	Urban	3	Non-Appalachia	0.600	0.880	25	13	46	15	5	66.3
21093	Hardin	Lincoln Trail	Urban	3	Non-Appalachia	0.644	0.854	261	136	45.2	168	70	73.1
21095	Harlan	Cumberland Valley	Rural	7	Appalachia	0.719	0.888	89	49	40.2	64	33	55.9
21097	Harrison	Bluegrass	Rural	6	Non-Appalachia	0.511	0.830	47	27	23.4	24	6	NA
21099	Hart	Barren River	Rural	8	Appalachia	0.617	0.883	60	41	36.3	37	20	54.2
21101	Henderson	Green River	Urban	2	Non-Appalachia	0.565	0.826	115	71	32.4	65	34	51.9
21103	Henry	Kipda	Urban	1	Non-Appalachia	0.595	0.881	42	28	22.8	25	13	80.7
21105	Hickman	Purchase	Rural	9	Non-Appalachia	0.429	0.786	14	13	20.5	6	6	30
21107	Hopkins	Pennyrile	Rural	4	Non-Appalachia	0.613	0.849	93	56	42.6	57	24	72.7
21109	Jackson	Cumberland Valley	Rural	9	Appalachia	0.548	0.839	31	19	27.1	17	9	63.1
21111	Jefferson	Kipda	Urban	1	Non-Appalachia	0.552	0.821	1957	1140	35.7	1080	470	69.4
21113	Jessamine	Bluegrass	Urban	2	Non-Appalachia	0.588	0.765	85	52	29	50	24	42.4
21115	Johnson	Big Sandy	Rural	7	Appalachia	0.438	0.672	64	42	29.5	28	13	47.9
21117	Kenton	Northern Kentucky	Urban	1	Non-Appalachia	0.618	0.845	283	170	34.4	175	81	65.2
21119	Knott	Kentucky River	Rural	9	Appalachia	0.569	0.843	51	35	18.6	29	15	50.4
21121	Knox	Cumberland Valley	Rural	7	Appalachia	0.544	0.842	57	33	35.9	31	15	52.6
21123	Larue	Lincoln Trail	Urban	3	Non-Appalachia	0.563	0.750	32	18	61.6	18	6	NA
21125	Laurel	Cumberland Valley	Rural	7	Appalachia	0.730	0.921	89	49	43	65	32	43.8
21127	Lawrence	Fivco	Rural	6	Appalachia	0.590	0.795	39	22	40.4	23	11	58.4
21129	Lee	Kentucky River	Rural	9	Appalachia	0.625	0.750	16	8	44.7	10	4	NA
21131	Leslie	Kentucky River	Rural	9	Appalachia	0.364	0.697	33	22	24.4	12	5	NA
21133	Letcher	Kentucky River	Rural	9	Appalachia	0.738	0.918	61	33	59.4	45	19	61.6
21135	Lewis	Buffalo Trace	Rural	8	Appalachia	0.690	0.929	42	26	30.7	29	16	30.8
21137	Lincoln	Bluegrass	Rural	7	Appalachia	0.721	0.930	43	24	54.7	31	16	62.1
21139	Livingston	Pennyrile	Rural	9	Non-Appalachia	0.528	0.861	36	28	26	19	13	47.2
21141	Logan	Barren River	Rural	6	Non-Appalachia	0.649	0.894	94	55	43.5	61	29	62.1
21143	Lyon	Pennyrile	Rural	8	Non-Appalachia	0.667	0.972	36	23	30.7	24	14	47.1

21145	McCracken	Purchase	Rural	5	Non-Appalachia	0.581	0.843	172	110	30.7	100	42	103
21147	McCreary	Lake Cumberland	Rural	9	Appalachia	0.589	0.768	56	39	25.3	33	19	44.8
21149	McLean	Green River	Urban	3	Non-Appalachia	0.703	0.919	37	21	42.6	26	13	60.7
21151	Madison	Bluegrass	Rural	4	Appalachia	0.639	0.819	166	102	27.5	106	54	40.2
21153	Magoffin	Big Sandy	Rural	9	Appalachia	0.656	0.844	32	20	23.9	21	11	34
21155	Marion	Lincoln Trail	Rural	6	Non-Appalachia	0.571	0.905	42	25	56.8	24	12	66.9
21157	Marshall	Purchase	Rural	7	Non-Appalachia	0.670	0.912	91	55	48.5	61	31	61.8
21159	Martin	Big Sandy	Rural	8	Appalachia	0.500	0.692	26	15	32.8	13	4	NA
21161	Mason	Buffalo Trace	Rural	6	Non-Appalachia	0.702	0.872	47	30	37.3	33	18	42.5
21163	Meade	Lincoln Trail	Urban	1	Non-Appalachia	0.513	0.795	78	52	39.9	40	19	65.6
21165	Menifee	Gateway	Rural	9	Appalachia	0.654	0.808	26	17	47.3	17	11	47.3
21167	Mercer	Bluegrass	Rural	6	Non-Appalachia	0.674	0.935	46	26	32.1	31	16	38.6
21169	Metcalfe	Barren River	Rural	9	Appalachia	0.667	0.792	24	12	70.4	16	6	70.4
21171	Monroe	Barren River	Rural	9	Appalachia	0.605	0.868	38	24	36.6	23	13	45.6
21173	Montgomery	Gateway	Rural	6	Appalachia	0.736	0.887	53	32	46.7	39	20	65.5
21175	Morgan	Gateway	Rural	7	Appalachia	0.676	0.912	34	20	30.3	23	12	50.7
21177	Muhlenberg	Pennyrile	Rural	6	Non-Appalachia	0.684	0.842	95	61	37.5	65	36	54.7
21179	Nelson	Lincoln Trail	Urban	1	Non-Appalachia	0.549	0.814	102	55	48.3	56	24	89.8
21181	Nicholas	Bluegrass	Rural	8	Appalachia	0.696	0.870	23	13	42.5	16	9	43.7
21183	Ohio	Green River	Rural	6	Non-Appalachia	0.614	0.904	83	55	26.7	51	30	51.5
21185	Oldham	Kipda	Urban	1	Non-Appalachia	0.714	0.929	98	57	45.7	70	34	61.1
21187	Owen	Northern Kentucky	Rural	8	Non-Appalachia	0.636	0.864	22	15	27.1	14	7	41.1
21189	Owsley	Kentucky River	Rural	9	Appalachia	0.636	0.864	22	14	23.1	14	7	45.6
21191	Pendleton	Northern Kentucky	Urban	1	Non-Appalachia	0.618	0.912	34	21	31.4	21	13	30.4
21193	Perry	Kentucky River	Rural	7	Appalachia	0.505	0.825	103	74	25.3	52	29	40.8
21195	Pike	Big Sandy	Rural	7	Appalachia	0.684	0.809	225	152	32.3	154	91	46.2
21197	Powell	Bluegrass	Rural	6	Appalachia	0.660	0.860	50	27	34.9	33	17	37.9
21199	Pulaski	Lake Cumberland	Rural	5	Appalachia	0.614	0.850	153	93	33.2	94	45	60.7
21201	Robertson	Buffalo Trace	Rural	8	Appalachia	0.800	1.000	5	3	46.8	4	2	54.5
21203	Rockcastle	Cumberland Valley	Rural	7	Appalachia	0.725	0.825	40	22	42.2	29	13	61.6
21205	Rowan	Gateway	Rural	7	Appalachia	0.660	0.830	47	30	32.5	31	17	56.8
21207	Russell	Lake Cumberland	Rural	9	Appalachia	0.524	0.786	42	26	27.4	22	9	NA
21209	Scott	Bluegrass	Urban	2	Non-Appalachia	0.683	0.921	63	35	39.7	43	18	69.2
21211	Shelby	Kipda	Urban	1	Non-Appalachia	0.629	0.835	97	49	56.2	61	20	115.7
21213	Simpson	Barren River	Rural	6	Non-Appalachia	0.745	0.836	55	36	34.2	41	22	53.7
21215	Spencer	Kipda	Urban	1	Non-Appalachia	0.586	0.862	29	12	NA	17	4	NA
21217	Taylor	Lake Cumberland	Rural	7	Non-Appalachia	0.535	0.733	101	61	34.7	54	26	51.6
21219	Todd	Pennyrile	Rural	8	Non-Appalachia	0.455	0.773	22	11	63.2	10	5	36.1
21221	Trigg	Pennyrile	Urban	3	Non-Appalachia	0.511	0.756	45	30	42	23	11	78.8
21223	Trimble	Kipda	Urban	1	Non-Appalachia	0.632	0.895	19	13	40.6	12	8	49.6
21225	Union	Green River	Rural	6	Non-Appalachia	0.550	0.825	40	27	45.4	22	12	67.6
21227	Warren	Barren River	Urban	3	Non-Appalachia	0.640	0.860	278	168	31.3	178	85	63.5
21229	Washington	Lincoln Trail	Rural	8	Non-Appalachia	0.714	0.821	28	17	22.5	20	10	37.1
21231	Wayne	Lake Cumberland	Rural	7	Appalachia	0.638	0.851	47	28	35.6	30	14	46.2
21233	Webster	Green River	Urban	2	Non-Appalachia	0.643	0.857	28	20	25.9	18	11	28.7
21235	Whitley	Cumberland Valley	Rural	7	Appalachia	0.632	0.832	95	57	30.5	60	31	51.4
21237	Wolfe	Kentucky River	Rural	9	Appalachia	0.625	0.792	24	15	22.5	15	8	49.1
21239	Woodford	Bluegrass	Urban	2	Non-Appalachia	0.586	0.845	58	31	26.6	34	15	79.8

R Script for data analysis

```
# Jeremiah Martin 10719733
# Capstone Project
# Analysis of KCR data
# Lung Cancer (NSCLC) - early stage - paterns/trends in definitive surgery


# Open File
# Macbook Pro
load(file="/Users/jerrymartin/Dropbox/Academics/000 Masters/CPH 778
Capstone/Data/kcr_nscl_0514.rda")
# Macbook Pro - Parallels / Windows
load(file="Z:/Dropbox/Academics/000 Masters/CPH 778
Capstone/Data/kcr_nscl_0514.rda")
# Home PC
load(file="C:/Users/Jerry/Dropbox/Academics/000 Masters/CPH 778
Capstone/Data/kcr_nscl_0514.rda")
# Work PC
load(file="C:/Users/MartinJT/Dropbox/Academics/000 Masters/CPH 778
Capstone/Data/kcr_nscl_0514.rda")
# Load Libraries

library(gmodels)
library(tableone)
library(MASS)
library(rattle)
library(survival)
library(choroplethr)
library(ggplot2)

# Copy the loaded dataset to a new table, n1
n1 <- lung

####### Data and Variable Cleanup #########

#  [1] Create Analytic Groups for this project
#
# Create the "clinically operable group"
#This uses best available clinical data
#giving the "benefit of the doubt"
#Starts by assuming patient is operable
#Then: Does the patient have mets? Mediastinal nodes? Unresectable T?
n1$ClinOp <- 1
n1$ClinOp[!(n1$TNMClinM %in% c("c0","cx=X"))] <- 0
n1$ClinOp[n1$TNMClinN %in% c("c2","c2A","c2C","c3","c3A")] <- 0
n1$ClinOp[n1$TNMClinT %in% c("c4","c4A")] <- 0


# create a categorical variable form BestStageGrp to categorize less than III
# This is "all comers" best stage divided into potentially operable vs not
n1$BestStageClass <- factor(ifelse(n1$BestStageGrp %in% c("00", "12", "15", "32",
"33"),1,0))

# Look at the "First Treatment Composite Code" and create to dummy classification
variables
# First is surgery of the primary site, ideal
# Second is surgery or radiation of the primary site - any focused local therapy

n1$surg <- factor(ifelse(n1$FstTrtCompCode %in% c(1,3,5,7,9,11,13,15),1,0))
n1$local <- factor(ifelse(n1$FstTrtCompCode %in%
c(1,3,4,5,6,7,9,11,12,13,14,15),1,0))

#   [2] Clean up, categorize the input variables
# Convert Race to a binary variable (overwhelming majority are white)
n1$rwhite <- factor(ifelse(n1$Race1 == "1", 1,0))

#convert tumor size to numeric and recode
n1$CSTumorSize <- as.numeric(n1$CSTumorSize)

#Collapse case class (diagnosed at treating facility vs referred in)
n1$caseclass.ref <- factor(ifelse(n1$CASECLASS %in%
c(0,10,11,12,13,14,20,21,22),0,1))

#collapse histologic diagnosis field
n1$histconf <- factor(ifelse(n1$DIAGCONFIRM %in% c(1,2,3),1,0))

#recode laterality where unspecified to NA
n1$Laterality[n1$Laterality>2] <- NA
n1$Laterality[n1$Laterality==0] <- NA
n1$Laterality <- factor(n1$Laterality)

#recode TUMOR_SIZE to missing where appropriate (>990)
n1$TUMOR_SIZE <- as.numeric(n1$CSTumorSize)
n1$TUMOR_SIZE[n1$TUMOR_SIZE>899] <- NA

#recode Marital Status of "unknown" to NA
n1$MARITALSTATUS[n1$MARITALSTATUS==9] <- NA
n1$MARITALSTATUS <- factor(n1$MARITALSTATUS)

#collapse insurance status
n1$insurance[n1$PRIMARYPAYOR %in% c(1,2,10,99)] <- 0 #unknown, self-pay,
uninsured
n1$insurance[n1$PRIMARYPAYOR %in% c(20,21)] <- 1 #managed care, private insurance
n1$insurance[n1$PRIMARYPAYOR %in% c(31,35)] <- 2 #medicaid
n1$insurance[n1$PRIMARYPAYOR %in% c(60,61,62,63,64)] <- 3 #medicare
n1$insurance[n1$PRIMARYPAYOR %in% c(65,66,67,68)] <- 4 #military, va, other govt
n1$insurance <- factor(n1$insurance)

#collapse smoking status
n1$smoking[n1$Tobacco %in% c(1,2,3,4)] <- 1 # smoker, chew, other kinds
n1$smoking[n1$Tobacco == 0] <- 0 # no smoking
n1$smoking[n1$Tobacco == 9] <- NA # unknown
n1$smoking <- factor(n1$smoking)

#create county beale code variable
n1$beale[n1$county %in%
c(21015,21023,21029,21037,21077,21081,21103,21111,21117,21163,21179,21185,21191,2
1211,21215,21223)] <- 1
n1$beale[n1$county %in%
c(21015,21017,21019,21049,21067,21089,21101,21113,21209,21233,21239)] <- 2
n1$beale[n1$county %in%
c(21015,21047,21059,21061,21091,21093,21123,21149,21221,21227)] <- 3
n1$beale[n1$county %in% c(21015,21073,21107,21151)] <- 4
n1$beale[n1$county %in% c(21015,21145,21199)] <- 5
n1$beale[n1$county %in%
c(21015,21003,21005,21009,21033,21041,21043,21055,21065,21079,21085,21097,21127,2
1141,21155,21161,21167,21173,21177,21183,21197,21213,21225)] <- 6
n1$beale[n1$county %in%
c(21015,21001,21013,21021,21025,21035,21051,21069,21071,21075,21083,21095,21115,2
1121,21125,21137,21157,21175,21193,21195,21203,21205,21217,21231,21235)] <- 7
n1$beale[n1$county %in%
c(21015,21011,21027,21031,21087,21099,21135,21143,21159,21181,21187,21201,21219,2
1229)] <- 8
n1$beale[n1$county %in%
c(21015,21007,21039,21045,21053,21057,21063,21105,21109,21119,21129,21131,21133,2
1139,21147,21153,21165,21169,21171,21189,21207,21237)] <- 9
```

```r
n1$beale <- factor(n1$beale)

#create county rate of surgery variable
n1$countyrate[n1$county %in% c(21131, 21063, 21105, 21115, 21219, 21039, 21001,
21013, 21087, 21005, 21159, 21193, 21047, 21097, 21221, 21163, 21207, 21043,
21139)] <- 1
n1$countyrate[n1$county %in% c(21217, 21023, 21071, 21121, 21109, 21179, 21073,
21225, 21111, 21025, 21027, 21123, 21021, 21101, 21119, 21069, 21017, 21155,
21033)] <- 2
n1$countyrate[n1$county %in% c(21015, 21145, 21037, 21045, 21239, 21215, 21113,
21089, 21147, 21127, 21079, 21103, 21091, 21171)] <- 3
n1$countyrate[n1$county %in% c(21035, 21059, 21051, 21107, 21199, 21183, 21099,
21191, 21117, 21081, 21065, 21129, 21237, 21007, 21211, 21235, 21223)] <- 4
n1$countyrate[n1$county %in% c(21187, 21189, 21231, 21151, 21227, 21233, 21093,
21049, 21055, 21019, 21041, 21141, 21067, 21165, 21153, 21011, 21205)] <- 5
n1$countyrate[n1$county %in% c(21197, 21003, 21029, 21075, 21169, 21143, 21157,
21167, 21175, 21057, 21209, 21061, 21177, 21195, 21083, 21135, 21181)] <- 6
n1$countyrate[n1$county %in% c(21085, 21161, 21149, 21229, 21185, 21053, 21095,
21137, 21203, 21125, 21173, 21133, 21213, 21031, 21009, 21077, 21201)] <- 7
n1$countyrate <- factor(n1$countyrate)

#create a rural variable from beale data
n1$rural <- factor(ifelse(n1$beale %in% c(1,2,3),0,1))

#convert remaining variables of interest to factors
n1$Sex <- factor(n1$Sex)
n1$Year_Diag <- factor(n1$Year_Diag)
n1$county <- factor(n1$county)
n1$APPAL <- factor(n1$APPAL)


#   [3] Prepare survival data
# Calculate survival dates:
# subtract 5-digit day code for date of last contact from date of diagnosis
# Create a survival in months variable

n1$survdays <- n1$Date_LC-n1$Date_dx
n1$survmos <- n1$survdays/30.4167


#Create new datasets for analysis
#lc = clinically operable lung group from ClinOp
#lb = best clinical stage operable group
lc <- subset(n1,n1$ClinOp == 1)
lb <- subset(n1,n1$BestStageClass == 1)


######## Create Table One ##########

#
# all variables of interest
# inputs <- c("MARITALSTATUS", "rwhite", "Sex", "DiagAge",
"CentralSequenceNumber",
#             "Laterality", "histconf", "caseclass.ref", "insurance",
#             "TNMClinT", "TNMClinN", "TNMClinM", "TNMClinStageGrp",
"TUMOR_SIZE",
#             "DerivedAJCC7StgGrp", "RXSummSurgPrimSite",
#             "REASONNOSURG", "ReasonNoRad", "VITALSTAT", "smoking",
"MenopauStatus",
#             "FstTrtCompCode", "BestStageGrp",
#             "Year_Diag", "county", "APPAL", "HS_Edu2010",
"PctBelowPoverty2010",
#             "ClinOp", "BestStageClass", "surg", "local")
#
#
# inputs.factor <- c("MARITALSTATUS", "rwhite", "Sex", "CentralSequenceNumber",
#                    "Laterality" , "histconf", "caseclass.ref", "insurance",
#                    "TNMClinT", "TNMClinN", "TNMClinM", "TNMClinStageGrp",
#                    "DerivedAJCC7StgGrp", "RXSummSurgPrimSite",
#                    "REASONNOSURG", "ReasonNoRad", "VITALSTAT", "smoking",
"MenopauStatus",
#                    "FstTrtCompCode", "BestStageGrp",
#                    "Year_Diag", "county", "APPAL",
#                    "ClinOp", "BestStageClass", "surg", "local")

# publication variables
inputs <- c("MARITALSTATUS", "rwhite", "Sex", "DiagAge", "CentralSequenceNumber",
            "Laterality", "histconf", "caseclass.ref", "insurance",
            "TNMClinT", "TNMClinN", "TNMClinM", "TNMClinStageGrp", "TUMOR_SIZE",
            "smoking", "BestStageGrp",
            "Year_Diag", "APPAL", "beale", "rural", "HS_Edu2010",
"PctBelowPoverty2010")


inputs.factor <- c("MARITALSTATUS", "rwhite", "Sex", "CentralSequenceNumber",
                   "Laterality" , "histconf", "caseclass.ref", "insurance",
                   "TNMClinT", "TNMClinN", "TNMClinM", "TNMClinStageGrp",
                   "smoking", "BestStageGrp",
                   "Year_Diag","APPAL", "beale", "rural")



###
### Create Descriptive Table 1
###

#Table of county by surgery or local therapy
#create maps using choroplethr

table(lc$county, lc$surg)
table(lc$county, lc$local)

# #need to load county rates for surgery and local therapy
# county <- data.frame(region = local_county$X1,value=local_county$X2)
# county_choropleth(county, "Local Therapy", state_zoom = "kentucky") +
scale_fill_brewer("Local Therapy",palette="Blues")
#



#Table 1a - Clinically "Benefit of doubt" best case, stratified by surgery vs
none
options(width=200)
t1 <- CreateTableOne(vars=inputs, strata="surg", data=lc, factorVars =
inputs.factor)
print(t1, missing=T, quote = T, showAllLevels=T, format="fp",
nonnormal=c("DiagAge", "CSTumorSize"))

#Table 1b - Best staging available, stratified by surgery vs none
t1 <- CreateTableOne(vars=inputs, strata="surg", data=lb, factorVars =
inputs.factor)
print(t1, missing=T, quote = T, showAllLevels=T, format="fp",
nonnormal=c("DiagAge", "CSTumorSize"))

#Table 1c - Clinically "Benefit of doubt" best case, stratified by local therapy
vs none
```

```r
t1 <- CreateTableOne(vars=inputs, strata="local", data=lc, factorVars =
inputs.factor)
print(t1, missing=T, quote = T, showAllLevels=T, format="fp",
nonnormal=c("DiagAge", "CSTumorSize"))

#Table 1d - Clinically "Benefit of doubt" best case, stratified by local therapy
vs none
t1 <- CreateTableOne(vars=inputs, strata="local", data=lb, factorVars =
inputs.factor)
print(t1, missing=T, quote = T, showAllLevels=T, format="fp",
nonnormal=c("DiagAge", "CSTumorSize"))


### Data Mining Techniques ####
# Using rattle() package
keepvars <- c("surg", "local", "MARITALSTATUS", "rwhite", "Sex", "DiagAge",
              "Laterality", "caseclass.ref", "insurance",
              "smoking", "county", "APPAL",
              "Year_Diag", "beale", "HS_Edu2010", "PctBelowPoverty2010")
lc.mine <- (lc[keepvars])
lb.mine <- (lb[keepvars])

rattle()

### Multivariate Modelling ########
# Logistic Reression model
# full <- glm(surg ~ MARITALSTATUS + rwhite + caseclass.ref + histconf +
TUMOR_SIZE
#              + insurance + smoking + Sex + DiagAge + CentralSequenceNumber +
Laterality
#              + APPAL + HS_Edu2010 + PctBelowPoverty2010, family="binomial",
data=lc)

# Usage: ldata is the dataframe of complete cases, replace the () with the source
data of interest

lkeep <- c("surg", "MARITALSTATUS", "rwhite", "Sex", "DiagAge",
              "Laterality", "caseclass.ref", "insurance",
              "smoking", "county", "APPAL",
              "Year_Diag", "beale", "HS_Edu2010", "PctBelowPoverty2010")

#Forward Selection using AIC (outcome surg)
ldata <- na.omit(lc[lkeep])
null <- glm(surg~1, data=ldata, family="binomial")
full <- glm(surg ~ ., data=ldata, family="binomial")
step (null, scope=list(lower=null, upper=full), direction="forward")


#This model is used to examine "local" as the outcome of interest
lkeepl <- c("local", "MARITALSTATUS", "rwhite", "Sex", "DiagAge",
           "Laterality", "caseclass.ref", "insurance",
           "smoking", "county", "APPAL",
           "Year_Diag", "beale", "HS_Edu2010", "PctBelowPoverty2010")

#Forward Selection using AIC (outcome local)
ldatal <- na.omit(lc[lkeepl])
null <- glm(local~1, data=ldatal, family="binomial")
full <- glm(local ~ ., data=ldatal, family="binomial")
step (null, scope=list(lower=null, upper=full), direction="forward")

#Copy here the final stepwise model selected from above
mylogit <-   glm(formula = surg ~ DiagAge + MARITALSTATUS + insurance +
caseclass.ref +
                   Year_Diag + smoking + Sex + rwhite + Laterality +
PctBelowPoverty2010,
                 family = "binomial", data = ldata)

mylogit_local <- glm(formula = local ~ DiagAge + MARITALSTATUS + caseclass.ref +
                       insurance + Sex + Year_Diag + PctBelowPoverty2010 + rwhite
+
                       beale, family = "binomial", data = ldatal)

summary(mylogit)
exp(cbind(OR = coef(mylogit), confint.default(mylogit)))

summary(mylogit_local)
exp(cbind(OR = coef(mylogit_local), confint.default(mylogit_local)))

# ROC CURVES


probs=predict(mylogit,lc,type=c("response"))
lc$probs=probs
library(pROC)
g <- roc(surg ~ probs, data = lc, plot=T)
g

probl=predict(mylogit_local,lc,type=c("response"))
lc$probl=probl
library(pROC)
g <- roc(local ~ probl, data = lc, plot=T)
g


#univariate odds ratios
unimylogit <-   glm(formula = surg ~ APPAL ,
                 family = "binomial", data = ldata)
summary(unimylogit)
exp(cbind(OR = coef(unimylogit), confint.default(unimylogit)))



####### Survival Analysis #########
lung.surv <- Surv(lc$survmos,lc$VITALSTAT==0) ~ lc$surg
surv.fit <- survfit(lung.surv, conf.type="none")
survdiff(lung.surv)
plot(surv.fit, xmax=72, main="Kaplan-Meier estimates for Surgery vs None",
     sub="Clinical best-case group",
     xlab="Time (Months)", ylab="Survival", lwd=2, lty=2:1, cex=0)
legend(x="topright", lwd=2, lty=1:2, legend=c("Surgery","No Surgery"))
text(10,0.1, labels="Log-Rank p<0.001")
print(surv.fit)


lung.surv <- Surv(lb$survmos,lb$VITALSTAT==0) ~ lb$surg
surv.fit <- survfit(lung.surv, conf.type="none")
survdiff(lung.surv)
plot(surv.fit, xmax=72, main="Kaplan-Meier estimates for Surgery vs None",
     sub="Best available stage group",
     xlab="Time (Months)", ylab="Survival", lwd=2, lty=2:1, cex=0)
legend(x="topright", lwd=2, lty=1:2, legend=c("Surgery","No Surgery"))
text(10,0.1, labels="Log-Rank p<0.001")
print(surv.fit)

lung.surv <- Surv(lc$survmos,lc$VITALSTAT==0) ~ lc$local
```

```
surv.fit <- survfit(lung.surv, conf.type="none")
survdiff(lung.surv)
plot(surv.fit, xmax=72, main="Kaplan-Meier estimates for Local Therapy vs None",
     sub="Clinical best-case group",
     xlab="Time (Months)", ylab="Survival", lwd=2, lty=2:1, cex=0)
legend(x="topright", lwd=2, lty=1:2, legend=c("Local Therapy","No Local
Therapy"))
text(10,0.1, labels="Log-Rank p<0.001")
print(surv.fit)


lung.surv <- Surv(lb$survmos,lb$VITALSTAT==0) ~ lb$local
surv.fit <- survfit(lung.surv, conf.type="none")
survdiff(lung.surv)
plot(surv.fit, xmax=72, main="Kaplan-Meier estimates for Local Therapy vs None",
     sub="Best available stage group",
     xlab="Time (Months)", ylab="Survival", lwd=2, lty=2:1, cex=0)
legend(x="topright", lwd=2, lty=1:2, legend=c("Local Therapy","No Local
Therapy"))
text(10,0.1, labels="Log-Rank p<0.001")
print(surv.fit)

#univariate hazard ratios
skeep <- c("surg","local", "MARITALSTATUS", "rwhite", "Sex", "DiagAge",
           "Laterality", "caseclass.ref", "insurance",
           "county", "APPAL", "countyrate",
           "Year_Diag", "beale", "HS_Edu2010", "PctBelowPoverty2010",
"VITALSTAT", "survmos")
ls <- lc[skeep] #which dataset lc vs lb to use in the following computations

summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$surg))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$local))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$beale))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$rural))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$APPAL))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$Sex))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$DiagAge))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$smoking))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$Year_Diag))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$HS_Edu2010))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$PctBelowPoverty2010))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$rwhite))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$caseclass.ref))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$insurance))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$countyrate))
summary(coxph(Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$MARITALSTATUS))


lung.surv2 <- Surv(ls$survmos,ls$VITALSTAT==0) ~ ls$surg+ ls$APPAL  + ls$Sex +
ls$DiagAge + ls$HS_Edu2010 + ls$PctBelowPoverty2010 + ls$rwhite +
ls$caseclass.ref + ls$insurance + ls$MARITALSTATUS
coxph.fit <- coxph(lung.surv2)
summary(coxph.fit)
```

EXEMPTION CERTIFICATION

MEMO: Jeremiah Martin, MB,BCh,FRC
3311 Old Post Road,
Portsmouth, OH, 45662
PI phone #: (203)809-6847

FROM: Institutional Review Board
c/o Office of Research Integrity

SUBJECT: Exemption Certification for Protocol No. 16-0892-X1B

DATE: October 22, 2016

On October 21, 2016, it was determined that your project entitled, *Disparities in Stage-Appropriate Therapy for Resectable Non-Small Cell Lung Cancer in Kentucky,* meets federal criteria to qualify as an exempt study.

Because the study has been certified as exempt, you will not be required to complete continuation or final review reports. However, it is your responsibility to notify the IRB prior to making any changes to the study. Please note that changes made to an exempt protocol may disqualify it from exempt status and may require an expedited or full review.

The Office of Research Integrity will hold your exemption application for six years. Before the end of the sixth year, you will be notified that your file will be closed and the application destroyed. If your project is still ongoing, you will need to contact the Office of Research Integrity upon receipt of that letter and follow the instructions for completing a new exemption application. It is, therefore, important that you keep your address current with the Office of Research Integrity.

For information describing investigator responsibilities after obtaining IRB approval, download and read the document "PI Guidance to Responsibilities, Qualifications, Records and Documentation of Human Subjects Research" from the Office of Research Integrity's IRB Survival Handbook web page [http://www.research.uky.edu/ori/IRB-Survival-Handbook.html#PIresponsibilities]. Additional information regarding IRB review, federal regulations, and institutional policies may be found through ORI's web site [http://www.research.uke.edu/ori]. If you have questions, need additional information, or would like a paper copy of the above mentioned document, contact the Office of Research Integrity at (859) 257-9428.

VITA

Dr. Jeremiah Martin is a cardiothoracic surgeon, originally from Ireland, now practicing at Southern Ohio Medical Center in Portsmouth Ohio. He undertook this Masters of Science in Clinical Trial Design (MSCRD) under the direction of Dr. David Mannino while working as an assistant professor in the department of Cardiothoracic Surgery at UK. His interest is in increasing awareness of early stage lung cancer, improving screening, and optimizing access to appropriate stage-directed therapy. He can be reached at martinjt@somc.org.