2012

Comprehensive Model of Lung Cancer Prediction and Prognosis

Joseph J. Putila
West Virginia University

Comprehensive Model of Lung Cancer Prediction and Prognosis

Joseph J. Putila

Dissertation submitted to the

School of Medicine

At West Virginia University

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

In

Public Health Sciences

Lan Guo, Ph.D., Chair

James Denvir, Ph.D.

Alan Ducatman, M.D., MSc

Sarah Knox, Ph.D.

Anoop Shankar, Ph.D.

Morgantown, West Virginia

2012

Keywords: Cancer, Lung, Prediction, Prognosis, Survival, Cox

Abstract

Comprehensive Model of Lung Cancer Prediction and Prognosis

Joseph J. Putila

Two unresolved issues in the treatment of non-small cell lung cancer are the assessment of risk of recurrence beyond the use of tumor stage alone, and the selection of an effective chemotherapeutic agent for patients with similar tumor morphology. A prognostic model able to identify high or low-risk patients with a high degree of accuracy can be used to inform clinicians on potential improvements to the current clinical practice. Clinical presentation, pathology, demographics, and genomics have been independently verified as influencing survival. A comprehensive model capable of incorporating multiple predictors into a unified measure has the potential to simplify risk assessment and more accurately model the determinants of patient outcome.

In order to accomplish this, patient characteristics including tumor stage, grade, patient race, age, COPD status ,and sex were assessed using Cox proportional hazards modeling across combinations of surgical, radiological, and chemotherapeutic treatments. A comprehensive model combining these factors was created and showed superior prognostic ability when compared to stage alone. In order to identify miRNA markers for chemoresponse, this patient data was then compared with information on miRNA expression from both a clinical cohort and the NCI-60 anti-cancer screen. A set of predictive and prognostic miRNA were selected by measuring the association between miRNA expression and disease-specific patient survival. The sets of significant miRNA were seen to have strong associations with mechanisms of apoptosis and cell-cycle control in an analysis of networked molecules.

The results show that a comprehensive model lends itself to a more accurate assessment of patient risk, and that these improvements persist across a variety of patient profiles and treatment modalities. Additionally, miRNA expression appears to play a role in patient response to chemotherapy when assessed across categories of disease progression. Multiple miRNA showed significant associations with disease-specific survival in the population analysis. These associations were able to be corroborated in the clinical and cellular data, demonstrating that this approach may be useful for identifying broad patterns of genomic expression which influence sensitivity and resistance to chemotherapy, and hold promise in further developing clinical tools for prediction. It was shown that the large, well-annotated, and diverse patient sample derived from registry and administrative data can be leveraged to approach two of the major unresolved issues in the treatment of non-small cell lung cancer.

Acknowledgements

We thank Dr. John Rogers, Dr. Manish Monga, Dr. Scot Remick, and Dr. Barbara Ducatman at West Virginia University for thoughtful discussion. We thank Dajie Luo at West Virginia University for developing a web application for clinical decision making based on the comprehensive model presented in this study. We are grateful to Pamela Moats and April Feathers for retrieving patient data from the Tumor Registry of the Mary Babb Randolph Cancer Center. I also thank the members of my committee, Dr. Jim Denvir, Dr. Alan Ducatman, Dr. Sarah Knox, Dr. Anoop Shankar, and my Chair, Dr. Lan Guo, for their support. This work was supported by the National Institutes of Health (R01LM009500, P20RR16440 and its ARRA Supplement to N.L. G., P2016477 for software license and training).

Table of Contents

Table of Tables

Table of Figures

Chapter1:
Introduction

Lung cancer represents a significant burden on the health and well-being of the U.S. population, and many other industrialized nations across the world. The low cure rate of lung cancer and the pervasiveness of its main causal factor, smoking, act together to produce a disease which is both deadly and widespread. While efforts to reduce the prevalence of smoking offer the most straightforward preventative solution, treatment of lung cancer in a clinical setting will remain a key component of managing the population disease burden for decades to come.

Lung and Bronchus cancer is currently the second most common cancer for both U.S. males and females behind cancers of the prostate and breast, respectively, with nearly half as many cases (American Cancer Society, 2012). Despite being less prevalent than other sites, lung cancer manages to lead all other cancers in mortality for both men and women. Lung cancer is responsible for as many deaths per year as cancers of the prostate, colon, and pancreas combined in men. In women, the annual number of deaths from lung cancer is equal to the total for the next two deadliest cancers, breast and colon. Because of this, improvements in lung cancer care represent an opportunity to significantly reduce the burden of disease on the population level.

Two unresolved issues in the treatment of non-small cell lung cancer are the assessment of risk of recurrence, and the selection of an effective chemotherapeutic agent for patients with similar tumor morphology. Relapse is the major cause of treatment failure and a major cause of death even among patients treated with surgical resection. Tumor stage alone has been unable to accurately select patients who may benefit from additional treatment such as adjuvant chemotherapy. Because of this, many early-stage patients experience recurrence even with surgical resection. Currently, 35–50% of stage I NSCLC patients will relapse within five years (Hoffman et al., 2000; Naruke et al., 1989). Although tumor stage is the strongest prognostic factor, other factors such as sex, age, race, comorbidities, and tumor grade contribute to survival as well (Clegg et al., 2002; Visbal et al., 2004; Orourke et al., 1987; Brown et al., 1996). We hypothesized that a prognostic model capable of combining multiple factors which influence survival would be able to provide refined prognostication when compared to the use of tumor stage alone.

An advantage of a comprehensive model of survival is the ability to aid in the selection of an appropriate therapy for each patient by providing an estimated measure of risk. The identification of high-risk patients who stand to benefit from additional treatment, more aggressive treatment, or enhanced follow-up is one of the key applications of a comprehensive prognostic model. Accurately incorporating multiple factors which influence survival has utility in selecting patients who may benefit from a modified treatment strategy as a result of increased or decreased risk relative to patients with similar clinical profiles. Currently, there are no models for non-small cell lung cancer prognosis in widespread clinical use which are able to incorporate multiple prognostic factors, but similar tools have been successful for breast cancers (Ravdin et al., 2001). Similar approaches to combining prognostic factors have shown promise in improving prognostication in lung cancer, but are restricted to limited treatment groups such as chemoradiation (Dehing-Oberije et al., 2008; Dehing-Oberije et al., 2009).

In order to evaluate the comprehensive prognostic model, individual factors were thoroughly assessed both as independent predictors in their ability to stratify patients into high and low-risk groups in a Kaplan-Meier analysis, and as coefficients in independent and controlled Cox regressions. The presented approach ensures that the model is able to function in a robust manner across a wide range of patient characteristics in order to increase its usefulness and generalizability to other patient populations. Although there is no standardized approach for building or evaluating prognostic models, the method of bootstrap estimation and internal validation on Harrell's C, Nagelkerke's R^2, ROC, and Brier score provide a robust analytical framework in-line with established theory (Schumacher et al., 2003; Steyerberg et al., 2001). These estimates, as well as the assessment of factors in univariate Cox models and on patient stratification, give a reasonable assessment of the utility of additional factors in improving prognostication.

The second unresolved issue, treatment prediction, is another area where of the selction of biomarkers associated with patient survival can benefit clinical decision-making. This is particularly true for response to chemotherapy, as the response to a given therapeutic agent can vary greatly between patients with seemingly similar clinical profiles and tumor morphology. Because of this, there is much to be gained by selecting patients who may benefit from the administration of chemotherapy and the selection of a specific agent with a higher probability of being effective. Avoiding the use of chemotherapy in patients highly unlikely to benefit from its administration also reduces the chance of unwanted side-effects or complications while ensuring that healthcare budgets are used in the manner which is most beneficial for a given population.

MiRNA biomarkers have the ability to enhance treatment selection in patients with similar tumor morphology. The use of miRNA has several advantages over other available biomarkers. It has been shown that miRNA are more stable in prepared tissue relative to mRNA (Jung et al., 2010; Mraz et al., 2009). This increased stability offers a significant advantage when attempting to measure expression in existing patient samples, most notably those prepared using formalin fixation (Xi et al., 2007). In addition, miRNA can be found in circulating plasma (Chen et al., 2012; Mitchell et al., 2008). The ability to collect miRNA non-invasivlely is a sizable advantage in the practical application of diagnostic and predictive tests over other biomarkers which may require biopsy or resected tissue. The inherent regulatory effect of miRNA on post-transcriptional regulation and signalling also represents an oppurtunity to focus on broad patterns of epigenetic regulation which are less well-represented with other markers (Iorio & Croce, 2012; Avraham & Yarden, 2012). This advantage is highlighted by the networks of regulatory targets and associated molecules generated from the sets of significant miRNA. Multiple aspects of chemoresponse such as the control of apoptosis and the cell-cycle show a high degree of interconnectivity with the predictive miRNA set and their associated molecules.

Numerous studies have shown that mRNA biomarkers can be used for predicting survival in lung cancer patients (Wan et al., 2012; Raponi et al., 2006; Chen et al., 2007; Beer et al., 2002). Similar studies have shown that miRNA can also be used in the diagnosis and prognosis of lung cancer (Wang et al., 2009; Rabinowits et al., 2009; Yu et al., 2008; Markou et al., 2008; Fabbri et al., 2007; Raponi et al., 2009). The development of a biomarker model of chemoreponse has the potential to both enhance response rates seen in patients treated with chemotherapy and to highlight the potential regulatory networks which influence patient survival and mediate chemoresponse. The main limitations of clinical studies are small sample size and non-random treatment assignment. Because the sample sizes are small and patient survival is influenced by variable treatment assignment it is difficult to assess the role of miRNA within treatment sub-groups.

An approach is presented for the use of registry data in the valdiation of miRNA markers found to be significant in a clinical population, and as a source of additional markers. This approach aims to circumvent the common issue of inadequate sample size imparted by cost and recruitment barriers. It was hypothesized that by linking miRNA expression in tumor samples to information on patient treatment and follow-up by measures of tumor progression, it would be possible to examine the role of miRNA in influencing survival and chemoresponse given a variety of treatment approaches. Using a clinical cohort (Raponi et al., 2009), an initial set of prognostic miRNA were selected by assessing the

association between expression and survival using Cox modeling and Kaplan-Meier stratification. These miRNA were then validated in the population cohort by estimating a linear regression between average miRNA expression and average survival for each unique combination of tumor T,N,M, and grade. Predictive miRNA were selected through the use of linear regression and validated using Cox modeling and Kaplan-Meier stratification in groups of patients having received a specific agent. In each case, miRNA expression was assessed using multiple measures of association in order maximize the utility of the set as classifiers (Pepe et al., 2004).

The significant predictive miRNA were then validated using data from cell-line experiments quantifying chemoresponse relative to miRNA expression. The use of cell-lines provides an avenue for simulating *ex-vivo* analyses seen to be useful in validating markers of chemoresponse in lung and other cancers (Vogt et al., 2002; Gallion et al., 2006; Chen et al., 2011; Coleman et al., 2004). Finally, Ingenuity Pathways Analysis was used to generate functional networks in which the set of significant miRNA and associated molecules were over-represented. These networks centered around functional aspects of cell proliferation and death, and provided insight into some of the potential regulatory networks through which miRNA may influence chemoresponse.

This document is divided into five chapters, each representing a component of a theoretical framework for developing a comprehensive population-based model of prognosis and prediction for non-small cell lung cancer. Chapter 2 outlines the issue of combining multiple prognostic factors to improve prognostic ability over a model using only tumor stage. Chapter 3 extends the analytical framework established in Chapter 2 by including data on chemotherapy and co-morbid conditions. These data were derived from administrative records, and allow for the assessment of the contribution of COPD to the prognostic ability of the comprehensive model and assessment of the accuracy of predictions in patients treated with chemotherapy. Having demonstrated a method for using demographic and clinical information to improve prognostication, Chapter 4 explores the role of miRNA in mediating chemoresponse. As with the previous chapters, the large and diverse population sample is leveraged to select sets of prognostic and predictive miRNA biomarkers. The final chapter, Chapter 5, concludes by highlighting the advantages of the analytical frameworks used and areas for further refinement.

Chapter2:
Combining Clinical, Pathological, and Demographic Factors Refines Personalized Prognosis of Lung Cancer: A Population-based Study

Published in PLoS One

Putila J, Remick SC, Guo NL, 2011 Combining Clinical, Pathological, and Demographic Factors Refines Prognosis of Lung Cancer: A Population-Based Study. PLoS ONE 6(2): e17493. doi:10.1371/journal.pone.0017493

Abstract

Background: In the treatment of lung cancer, an accurate estimation of patient clinical outcome is essential for choosing an appropriate course of therapy. It is important to develop a prognostic model by combining clinical, pathological and demographic factors for individualized clinical decision making.

Methodology: A total of 234,412 patients diagnosed with adenocarcinomas or squamous cell carcinomas of the lung or bronchus between 1988 and 2006 were retrieved from the SEER database to construct a prognostic model. SEER patients were randomly assigned to a training or test set. Two additional patient cohorts (*n* = 1,991) from multiple hospitals were also used as test sets. A patient stratification scheme was developed on the training set using a Cox proportional hazard model with tumor stage, tumor grade, patient age, gender, and race included in the model. The training model was used to predict the clinical outcome for each patient in the test sets.

Principal Findings: The comprehensive model consistently outperformed the model using stage alone on Harrell's C, Nagelkerke's R^2, and Brier Scores. The comprehensive model also performed better than stage alone when splitting the SEER patients into four treatment modalities. Two additional patient cohorts (*n* = 1,991) were also used as external validation, with the comprehensive model again outperforming the model using stage alone on the three metrics used.

Conclusion: These results demonstrate the feasibility of constructing a precise prognostic model combining multiple clinical, pathologic, and demographic factors. The comprehensive model significantly

improves individualized prognosis upon AJCC tumor staging and is robust across a range of treatment modalities, the spectrum of patient risk, and in novel patient cohorts.

Introduction

Lung cancer is one of the most aggressive cancer types and consistently the leading cause of cancer-related death in the United States for both men and women. There are around 215,000 new cases and 161,000 deaths annually (Jemal et al., 2008). Non-small cell lung cancer (NSCLC) accounts for about 80% of lung cancer cases. Although tumor stage is strongly predictive of survival in most cases, it does not explain the distinct variability in treatment outcome within patients of the same stage. Currently, surgery is the major treatment option for patients with stage I NSCLC. However, 35–50% of stage I NSCLC patients will relapse within five years (Hoffman et al., 2000; Naruke et al., 1989), which is the major cause of treatment failure, i.e. death from lung cancer. It remains an unsolved challenge for physicians to reliably identify patients at high risk for tumor recurrence as candidates for adjuvant chemotherapy.

Prognostic factors such as age, gender, and tumor grade, have been shown to be strongly associated with survival. Age is an established risk factor for the development of lung cancer and can also influence the type of treatment received either due to medical coverage or the existence of co-morbid conditions which preclude certain therapies (Brown et al., 1996; Orourke et al., 1987). Males diagnosed with lung cancer consistently experience poorer survival than do females (Visbal et al., 2004). This gender difference persisted even when controlling for other variables such as tumor stage, age at diagnosis, and treatment.

Race has also been shown to be a significant predictor of survival, with Asians and Pacific Islanders experiencing better survival in both prospective (Thatcher et al., 2005) and population-based studies (Clegg et al., 2002). While the disease mechanism and genetic background is not well characterized, the consistency of this finding is useful in terms of prognostication and treatment.

The emerging use of genetic markers may enable physicians to make treatment decisions based on the specific characteristics of individual patients and their tumors, instead of population statistics (Dalton & Friend, 2006). This study presents an alternative avenue to improve personalized prognosis of NSCLC by combining clinical, pathological, and demographic factors in a population-based study (n = 234,412). This comprehensive model was tested across a number of treatment modalities and blindly validated on multiple separate patient cohorts (n = 1,991). The comprehensive model achieved a

significant improvement in prognostication when compare with AJCC tumor staging system including cases converted to AJCC 7th Edition (American Joint Committee on Cancer, 2010). This patient stratification scheme could be integrated with future clinically-validated prognostic gene signatures for personalized prognosis of NSCLC.

Materials and Methods

Acquisition of Patient Cohorts

A cohort of patients diagnosed with lung cancer was retrieved from the Surveillance Epidemiology and End Results (SEER) database (SEER, 2010). The SEER database is an aggregate of registry data from specific geographic areas covering approximately 26 percent of the U.S. population, and contains clinical, demographic, treatment, and follow-up information for a variety of cancers. The requirements for inclusion in the study included a diagnosis of primary lung adenocarcinoma (ICD-O-3 8140 to 8380) or squamous cell carcinoma (ICD-O-3 8050 to 8080) between the years 1988 and 2006, as well as available data on tumor stage, tumor grade, race, age, gender, disease-specific survival, and treatment. Patients who were diagnosed via autopsy or death certificate, or had no valid survival data were excluded from the analysis. A total of 234,412 patients met the inclusion criteria. Patients staged using the 6th edition of AJCC staging, in general 2004 and newer diagnoses, were recoded to the 7th edition based on the proposed staging changes in the AJCC Staging Manual (American Joint Committee on Cancer, 2010) and information about tumor size, extension, metastasis, and lymph node involvement found in the SEER database where possible. A total of 58,634 cases were able to be converted from the 6th to the 7th edition.

Two additional patient cohorts were also used as validation sets. De-identified data for a total of 1,552 patients treated at the Mary Babb Randolph Cancer Center at West Virginia University from 1990 to 2009 with squamous cell carcinoma (*n* = 758) or adenocarcinoma (*n* = 794) were obtained. The study was approved with an IRB exemption from West Virginia University. According to HIPAA regulation, de-identified clinical information can be used in research without prior consent from the patients. A total of 439 lung adenocarcinoma cases were also obtained from Shedden et al (Shedden et al., 2008) for patients with Stage I-IIIB cancers. These patients were treated in H. Lee Moffitt Cancer Center, University of Michigan Comprehensive Cancer Center, Dana-Farber Cancer Institute, and Memorial Sloan-Kettering Cancer Center. Patients have provided consent. These data have been published in Shedden et al (Shedden et al., 2008) before. It is not clear if patients have provided written or verbal

consent. The protocols were approved with Institutional Review Boards (IRB-Med) of the respective institutes. The total numbers and distribution of patient characteristics are described in detail in Tables 2.1 and 2.2.

Table 2.1. Outline of patient clinical characteristics for major histology of non-small cell lung cancer and AJCC staging editions retrieved from SEER database.

	Adenocarcinoma		Squamous	
Variable	AJCC 3rd+6th	AJCC 7th	AJCC 3rd+6th	AJCC 7th
Age				
Mean Age (σ)	66.9 (11.4)	67.0 (11.3)	69.0 (9.9)	69.8 (10.2)
Sex				
Male (%)	75,753 (50.4)	18,550 (48.3)	55,794 (66.2)	12,678 (62.7)
Race				
API	10,377 (6.9)	2,853 (7.4)	3,877 (4.6)	885 (4.4)
Black	14,432 (9.6)	3,620 (9.4)	10,373 (12.3)	2,317 (11.5)
White	125,349 (83.5)	31,953 (83.2)	70,004 (83.1)	17,006 (84.2)
Tumor Stage*				
I	36,052 (24)	8,295 (21.6)	21,495 (25.5)	4,090 (20.2)
II	6,118 (4.1)	4,661 (12.1)	4,899 (5.8)	3,026 (15)
IIIA	11,447 (7.6)	5,773 (15)	11,284 (13.4)	4,497 (22.2)
IIIB	26,905 (17.9)	3,008 (7.8)	19,933 (23.7)	2,435 (12)
IV	69,636 (46.4)	16,690 (43.4)	26,643 (31.6)	6,160 (30.5)
Tumor Grade				
Grade 1	11,415 (7.6)	3,602 (9.4)	2,559 (3.0)	464 (2.3)
Grade 2	28,999 (19.3)	8,637 (22.4)	22,877 (27.2)	5,700 (28.2)
Grade 3	45,424 (30.3)	9,796 (25.5)	32,380 (38.4)	7,266 (36)
Grade 4	64,320 (42.8)	16,391 (42.7)	26,438 (31.4)	6,779 (33.5)

*Sub-stages for stage I and II patients are combined as it was not possible to differentiate between sub-stages for all patients diagnosed with AJCC 3rd and 6th staging systems. Categorical variables show the N and percentage of the total for each category. Age is represented as the mean age with the standard deviation in parentheses.

Table 2.2. Outline of patient clinical characteristics for external non-small cell lung cancer validation sets.

	Adenocarcinoma		Squamous
Variable	Director's Challenge (Shedden et al., 2008)	MBRCC	MBRCC
Age			
Mean Age (σ)	64.4 (10.1)	64.3 (11.3)	67.1 (10.1)
Sex			
Male	218 (50.3%)	419 (52.8)	479 (63.2)
Race			
API	7 (1.6)	2 (0.3)	1 (0.1)
Black	12 (2.7)	9 (1.1)	15 (2.0)
White	420 (95.7)	783 (98.6)	742 (97.9)
Tumor Stage			
I	276 (62.9)	181 (22.8)	176 (23.2)
II	95 (21.6)	48 (6)	57 (7.5)
IIIA	57 (13)	74 (9.3)	111 (14.6)
IIIB	11 (2.5)	95 (12)	115 (15.2)
IV	0 (0)	396 (49.9)	299 (39.4)
Tumor Grade			
Grade 1	60 (13.7)	62 (7.8)	22 (2.9)
Grade 2	208 (47.4)	137 (17.2)	172 (22.6)
Grade 3	166 (37.8)	231 (29)	267 (35.2)
Grade 4	5 (1.1)	364 (45.8)	297 (39.2)

Conversion of Cases to AJCC 7th Edition

Cases diagnosed from 2004 onward were able to be converted into the AJCC 7th Edition. The original TNM staging information regarding tumor size and extension (T), lymph node status (N), and distant metastasis (M) was retrieved from the SEER data. Using this information, the T, N, and M classifiers were recoded according to the new guidelines (American Joint Committee on Cancer, 2010) and then used to determine the AJCC 7th stage.

Model Construction and Statistical Analyses

Disease-specific survival was analyzed primarily using a Cox proportional hazards model. This model estimates the effect of a set of factors on the time until an event, in this case death, following a diagnosis. Four models, one for each of the histology and AJCC staging combinations, were estimated. A total of 500 bootstrapped samples equal in size to the original adenocarcinoma and squamous cell carcinoma patient cohorts were constructed. This method has been seen to be superior to split-sample techniques (Steyerberg et al., 2001), and in general produces less biased estimates with a smaller

variance. A Cox model was then fit on each bootstrapped sample. In order to determine the advantage of using other variables in addition to AJCC stage, two sets of variables were used. The first contained information on tumor stage and grade, patient age, race, and gender. The second contained only information on tumor stage and used as a model of current clinical practice. The final model used the mean value of all coefficients generated from the bootstrapped samples, as the distribution of hazard scores was normal. Hazard scores were calculated for each patient in the original samples based on the final model constructed from the means. The formula used to specify the model is shown below, demonstrating the relationship between hazard *h* for patient *i* at time *t* and the coefficients, β, for variables 1 through *k* with values of *x*.

$$\log h_i(t) = \log \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_i$$

Patients still alive or dead due to unrelated causes were censored at the time of last follow-up or death, respectively. Internal performance was measured using Harrell's C, Nagelkerke's R^2, and Brier Scores. Harrell's C is a measure of concordance which is representative of the area under an ROC curve ranging between 0 and 1, with higher scores indicating greater concordance (Steyerberg et al., 2001). Nagelkerke's R^2 is functionally similar to the R2 value in linear models, ranging between 0 and 1 with higher values explaining more variance, with this variant being calculated on the log-likelihood scale. The Brier score represents the average prediction error, ranging from 1 to 0, with lower values indicating a lower average error. The model constructed using the training set was then further validated on SEER sub-cohorts as well as patients from the MBRCC and the Director's Challenge (Shedden et al., 2008) cohorts, without re-estimating parameters of the model. Selection of patient stratification cutoffs was done by iterating through a range of possible cutoff values and measuring the difference in survival between High, Intermediate, and Low risk groups defined according to the Full and Stage Only models at each cutoff. The set of cutoff values that offered the greatest improvement in prognostication over the Stage Only model was selected for use in the final model. Statistical analyses were conducted with the *pamr*, *pec*, *Design,* and *survival* packages in *R* v2.11.0.

Results

This study focused on two major cell types of NSCLC, lung adenocarcinoma and squamous cell lung cancer. For each cell type, a comprehensive model was constructed to include the previous AJCC staging system (the 3rd and 6th editions) and the current AJCC 7th edition. The clinical characteristics of

the SEER patient population are listed in Table 2.1, and two external validation cohorts are summarized in Table 2.2.

The models were constructed by taking the mean of each coefficient from a Cox model fit on 500 bootstrapped samples of each original cohort. This resulted in a total of four models, one for each of the two AJCC staging systems and two cell types considered in combination with one another. These models were tested on the original samples in their entirety, sub-cohorts representative of four major treatment modalities, and two external cohorts.

In the overall studied patient population, earlier stage at diagnosis was significantly related to disease-specific survival in a univariate Cox Proportional Hazards model in both adenocarcinoma and squamous cell carcinoma for each AJCC Staging system ($P<0.05$). In the multivariate analyses AJCC stage, tumor grade, patient age, race, and gender were all significant. Specifically, lower tumor grade, younger age at diagnosis, and being of Asian/Pacific Islander descent were all significantly associated with improved survival ($P<0.05$). Being male or having a later stage at diagnosis was associated with a poorer outcome across all groups. The comprehensive model incorporating all these factors showed significantly improved prognostic categorization when compared with the AJCC staging system, including the latest edition. The improvement across all patients for each staging system and histology is demonstrated in Figure 2.1.

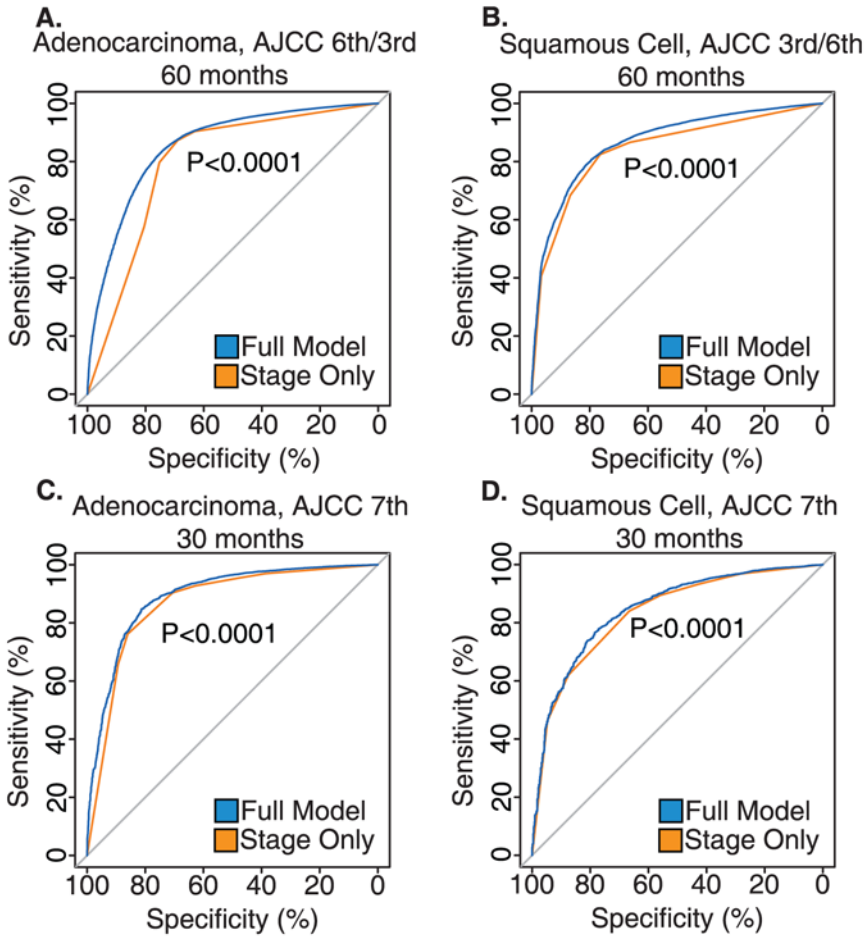

Figure 2.1. Prediction of survival at 60 months for the AJCC 3rd and 6th Editions (top) and 30 months for the cases converted to the AJCC 7th Edition (bottom) for both lung adenocarcinoma (left) and squamous cell lung cancer (right) using ROC curves. P<0.05 indicates that the full model is significantly more accurate in predicting disease-specific survival than tumor stage.

The patients were then assigned into one of four treatment categories based on the treatment record in SEER database. These categories were surgery alone, radiation alone, surgery with radiation, and no treatment listed. For simplicity, this determination was based on the presence or absence of any surgical or radiation procedure, regardless of the specific procedure.

Patient stratification for lung adenocarcinoma (the AJCC 3rd and 6th edition)

A total of 150,158 lung adenocarcinoma patients staged with the 3rd and 6th AJCC Editions met the criteria for inclusion. Harrell's c statistic was calculated for both the model using stage alone and the comprehensive model using additional variables. The comprehensive model had a higher C statistic

(0.732) compared to the stage only model (0.694). A similar improvement was seen for Nagelkerke's R^2 (0.294 v. 0.253) and Brier score (0.134 v. 0.143). The distribution of Hazard scores and general trend in improvement is shown in Figure 2.2 for all patients, and for patients treated with surgery alone.

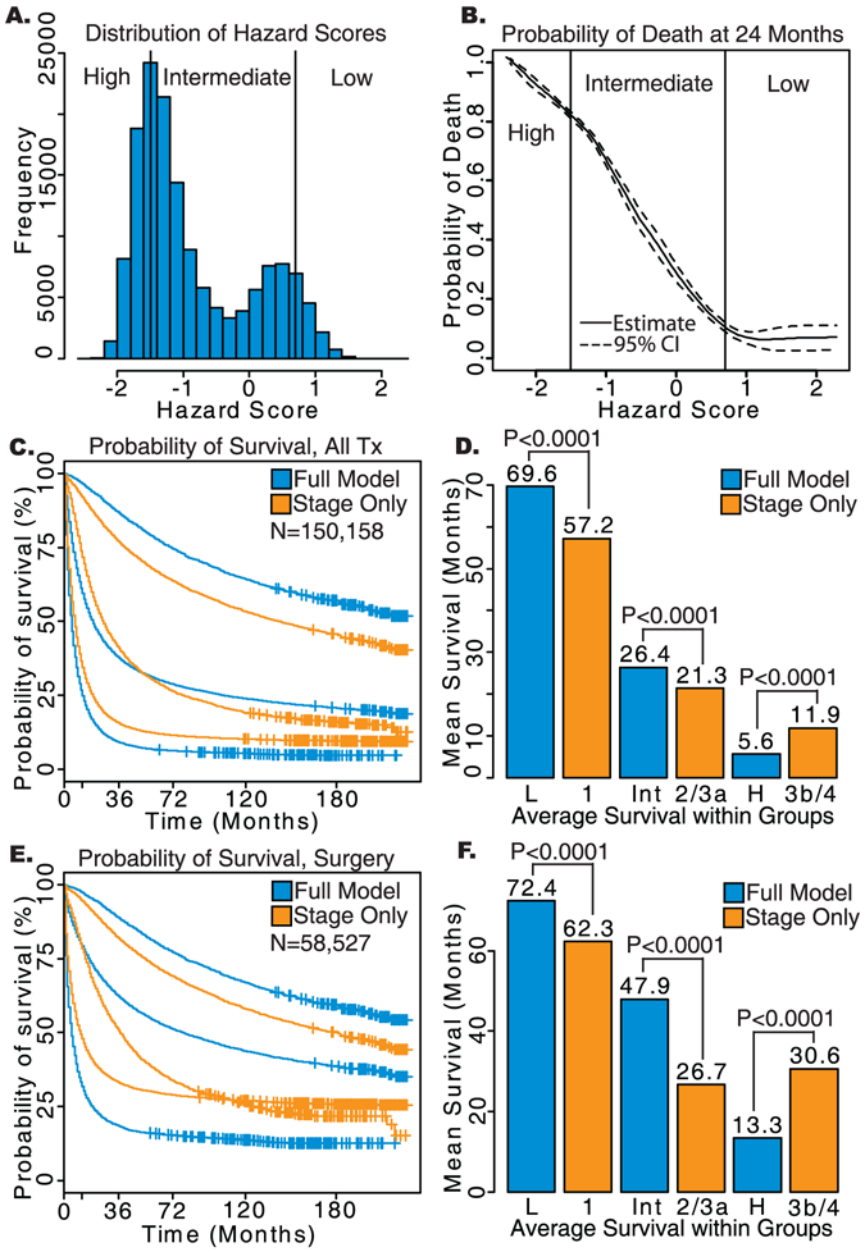

Figure 2.2. Results of survival analysis on lung adenocarcinoma patients staged using AJCC 3rd or 6th Edition. a) Histogram of Hazard Scores obtained from the comprehensive model. b) Probability of death from lung cancer prior to 24 months based on Hazard Scores calculated using the comprehensive model. c) Kaplan-Meier survival plots for low-, intermediate-, and high-risk groups determined by the comprehensive model (blue) and AJCC staging alone (orange). d) Average survival of each group in months, with log-rank P-values shown. L: low-risk; Int: intermediate-risk; H: high-risk defined by the full model. Stage only model contains patient with stage 1, 2, 3a, 3b and 4. e) Kaplan-Meier survival plots for each risk group in patients who received surgery without radiation. f) Average survival for risk groups in patients who received surgery without radiation. L: low-risk; Int: intermediate-risk; H: high-risk. Stage only model contains patient with stage 1, 2, 3a, 3b and 4.

The analysis comparing the performance of each model on treatment subgroups also showed a similar improvement in predictive ability with the comprehensive model. In patients who received surgery alone, the comprehensive model had consistently better estimates for Harrell's C (0.768 v 0.723), Nagelkerke's R^2 (0.225 v 0.173) and Brier Score (0.206 v. 0.210). A similar improvement, summarized in Tables 2.3, 2.4 and 2.5, was observed in patients receiving radiation alone, surgery with radiation, and those with no treatment listed.

Lung adenocarcinoma cases converted to the AJCC 7th edition

A total of 38,426 lung adenocarcinoma cases were converted into the AJCC 7th edition. It is important to note that the converted cases represent a much smaller cohort and have shorter follow-up time compared to the AJCC 3rd and 6th Edition cohorts. When considering the entire patient sample, Harrell's C for the comprehensive v. the stage only model (0.763 v. 0.731), Nagelkerke's R^2 (0.305 v. 0.274) and Brier score (0.144 v 0.150) were all improved. These effects persisted in when considering the four patient sub-cohorts defined by treatment modality, although the performance of both models was similarly decreased when compared to the original staging system. The patient sub-cohort with no treatment listed performed the worst on all three metrics. Figure 2.3 shows the distribution of Hazard scores and improvement over the use of stage alone when constructing risk groups.

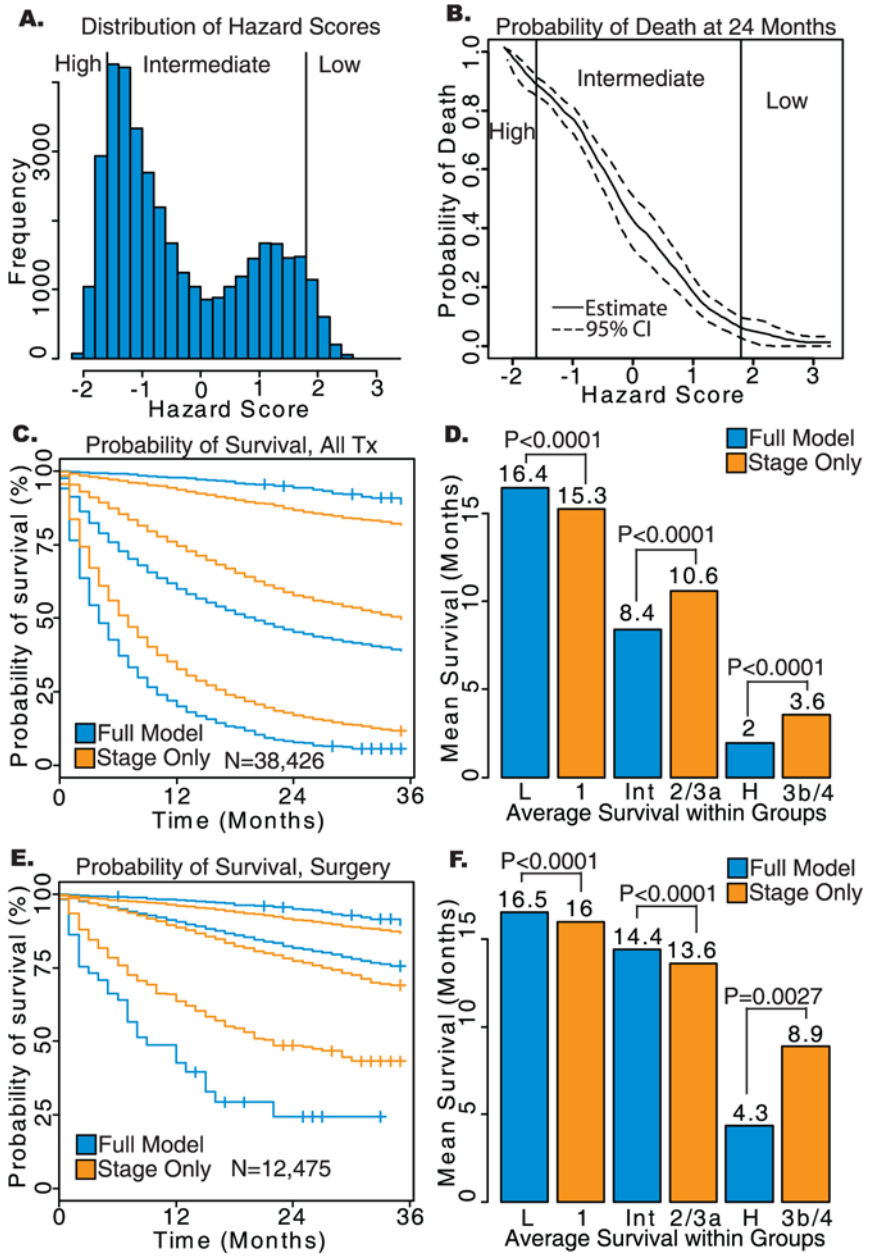

Figure 2.3. Results of survival analysis on lung adenocarcinoma patients converted to AJCC 7th Edition. a) Histogram of Hazard Scores obtained from the comprehensive model. b) Probability of death from lung cancer prior to 24 months based on Hazard Scores calculated using the comprehensive model. c) Kaplan-Meier survival plots for low-, intermediate-, and high-risk groups determined by the comprehensive model (blue) and AJCC staging alone (orange). d) Average survival of each group in months, with log-rank P-values shown. e) Kaplan-Meier survival plots for each risk group in patients who received surgery without radiation. f) Average survival for risk groups in patients who received surgery without radiation. L: low-risk; Int: intermediate-risk; H: high-risk defined by the full model. Stage only model contains patient with stage 1, 2, 3a, 3b and 4.

Prognostication of squamous cell lung cancer (the AJCC 3rd and 6th edition)

A total of 84,254 squamous cell lung cancer patients diagnosed with the ACC 3rd and 6th staging system met the inclusion criteria. Performance of both the comprehensive and stage only model were slightly decreased when compared to the adenocarcinoma patients in the overall patient sample. However, there was still an improvement in the overall treatment cohort when using the comprehensive model on Harrell's C (0.722 v. 0.706), Nagelkerke's R^2 (0.289 v. 0.274), but not on Brier score (0.119 v 0.119). There was a similar improvement in the sub-cohorts defined by treatment modality, with the comprehensive model performing as well or better than the stage only model in all sub-cohorts. Figure 2.4 shows the improvement in patients receiving any treatment, and those receiving surgery alone when constructing risk groups, as well as the distribution of Hazard scores.

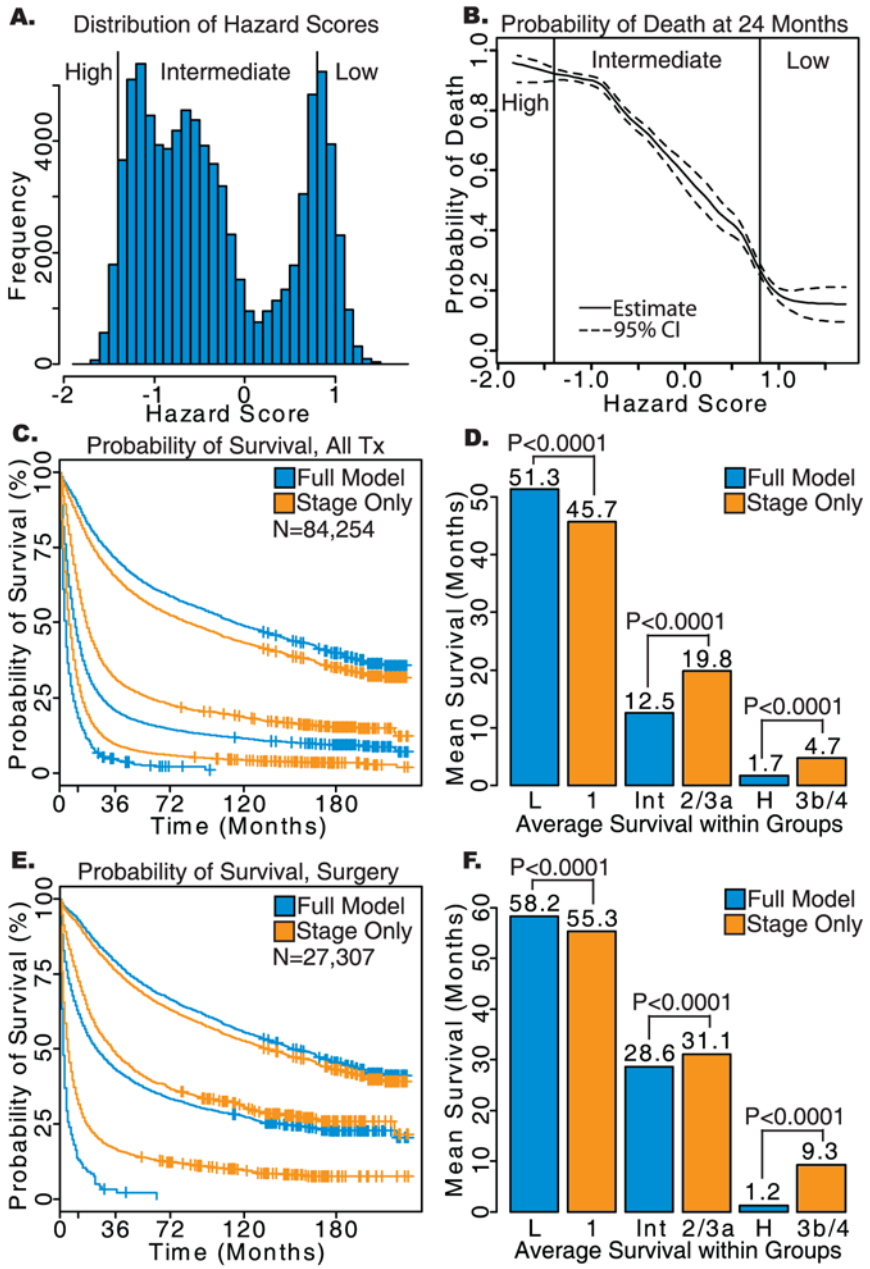

Figure 2.4. Results of survival analysis on squamous cell lung cancer patients staged using AJCC 3rd or 6th Edition. a) Histogram of Hazard Scores obtained from the comprehensive model. b) Probability of death from lung cancer prior to 24 months based on Hazard Scores calculated using the comprehensive model. c) Kaplan-Meier survival plots for low-, intermediate-, and high-risk groups determined by the comprehensive model (blue) and AJCC staging alone (orange). d) Average survival of each group in months, with log-rank P-values shown. e) Kaplan-Meier survival plots for each risk group in patients having received surgery without radiation. f.) Average survival for risk groups in patients who received surgery without radiation. L: low-risk; Int: intermediate-risk; H: high-risk defined by the full model. Stage only model contains patient with stage 1, 2, 3a, 3b and 4.

Squamous cell lung cancer cases converted to the AJCC 7th edition

A total of 20,208 squamous cell lung cancer cases could be converted to the AJCC 7th edition. Prediction was similar or improved when using the comprehensive model on all three metrics and in all treatment cohorts considered, however the difference between the two models was marginal in a few cases. The most marked improvement in prediction was in the sub-cohort of patients receiving surgery only. For that group, the comprehensive model outperformed the stage only model on Harrell's C (0.689 v. 0.670), Nagelkerke's R^2 (0.064 v. 0.055), and marginally on Brier score (0.113 v. 0.114). Figure 2.5 shows the improvement in risk group construction over the model using stage alone, as well as the distribution of Hazard scores.

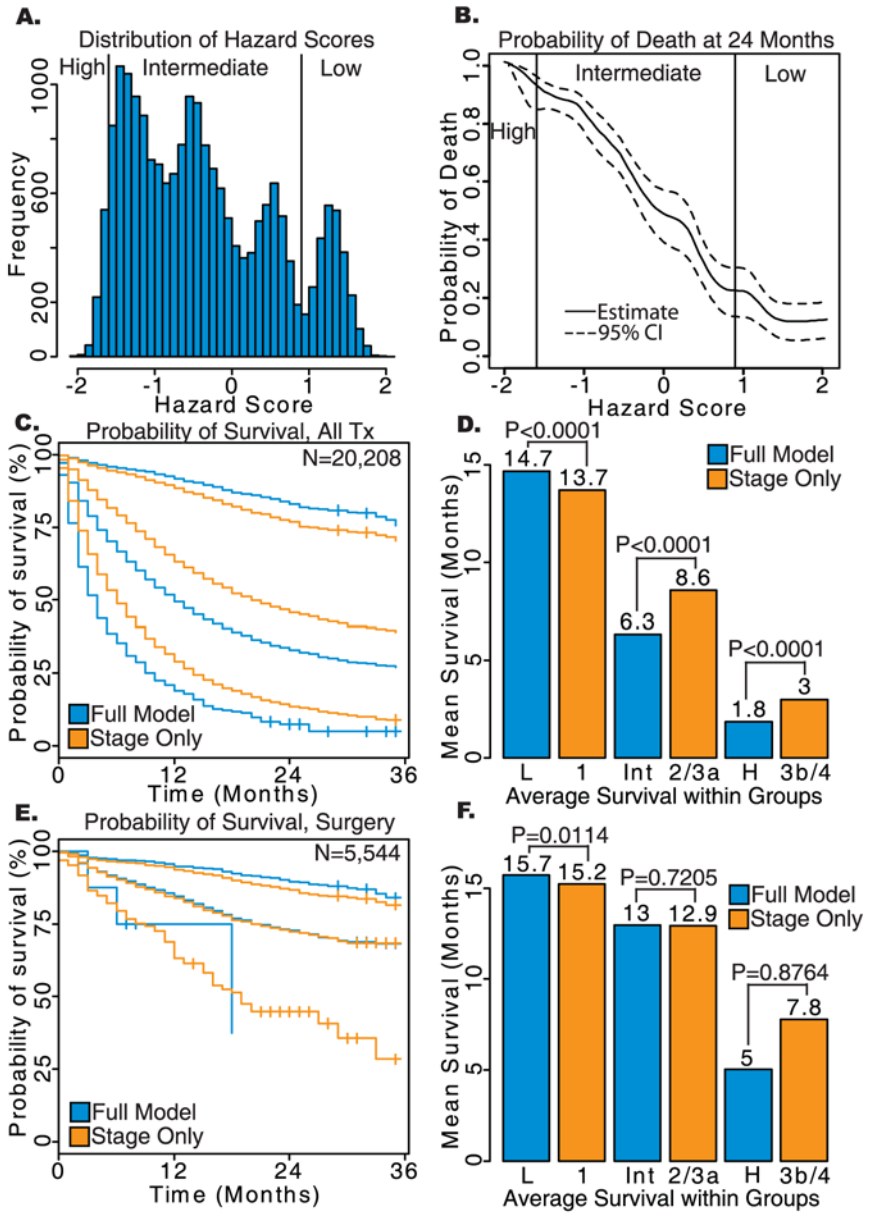

Figure 2.5. Results of survival analysis on squamous cell lung cancer patients converted to AJCC 7th Edition. a) Histogram of Hazard Scores obtained from the comprehensive model. b) Probability of death from lung cancer prior to 24 months based on Hazard Scores calculated using the comprehensive model. c) Kaplan-Meier survival plots for low-, intermediate-, and high-risk groups determined by the comprehensive model (blue) and AJCC staging alone (orange). d.) Average survival of each group in months, with log-rank P-values shown. e) Kaplan-Meier survival plots for each risk group in patients who received surgery without radiation. f.) Average survival for risk groups in patients who received surgery without radiation. L: low-risk; Int: intermediate-risk; H: high-risk defined by the full model. Stage only model contains patient with stage 1, 2, 3a, 3b and 4.

Treatment selection for stage I surgical patients

Patients with stage I cancers who were treated with surgery alone were extracted for further analysis to determine whether the comprehensive model could identify patients who may benefit from a more aggressive therapy. In most cases, stage I patients that received surgery had the most favorable outcome compared to other stages or treatment modalities, meaning that this cohort represents an extreme tail of the patient risk spectrum.

The models developed on the overall patient samples were applied to stage I patients that had received surgery alone in order to determine the predictive ability of the comprehensive and stage only models in a cohort with a more favorable outcome. For adenocarcinoma patients staged using AJCC 7th Edition (N=6,872), the comprehensive model performed better on Nagelkerke's R^2 (0.021 v. 0.002) and Brier score (0.0541 v. 0.0552) but did not outperform the stage only model on Harrell's C (0.371 v.0.469). However, when using the median hazard score to split patients into low and high-risk groups, the comprehensive model was able to select a group of low-risk patients with better survival (16.53 v. 16.15 months, P=0.0008). The comprehensive model performed similarly to the stage only model in squamous cell patients (N=2,953), with no significant difference between the risk groups defined by each model (P>0.05). The comprehensive model was also able to further separate stage IA and stage IB patient groups into high and low-risk groups with significantly different survival curves ($p<0.05$) for both adenocarcinoma and squamous cell carcinoma, as shown in Figures 2.6 and 2.7, respectively.

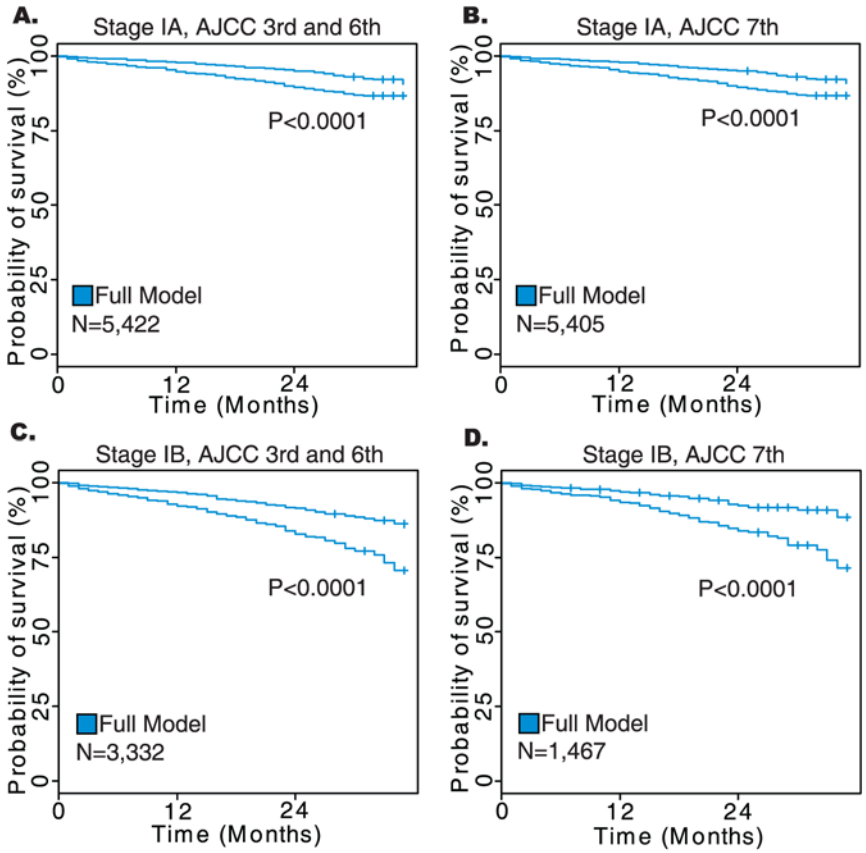

Figure 2.6. Results of survival analysis on lung adenocarcinoma patients diagnosed with stage IA or IB disease. The Kaplan-Meier plots show the difference between low- and high-risk groups as determined by the comprehensive model. Data on sub-stage was only available for patients staged using the AJCC 6th Edition staging system (2004 and later) and for those patients converted into the 7th Edition.

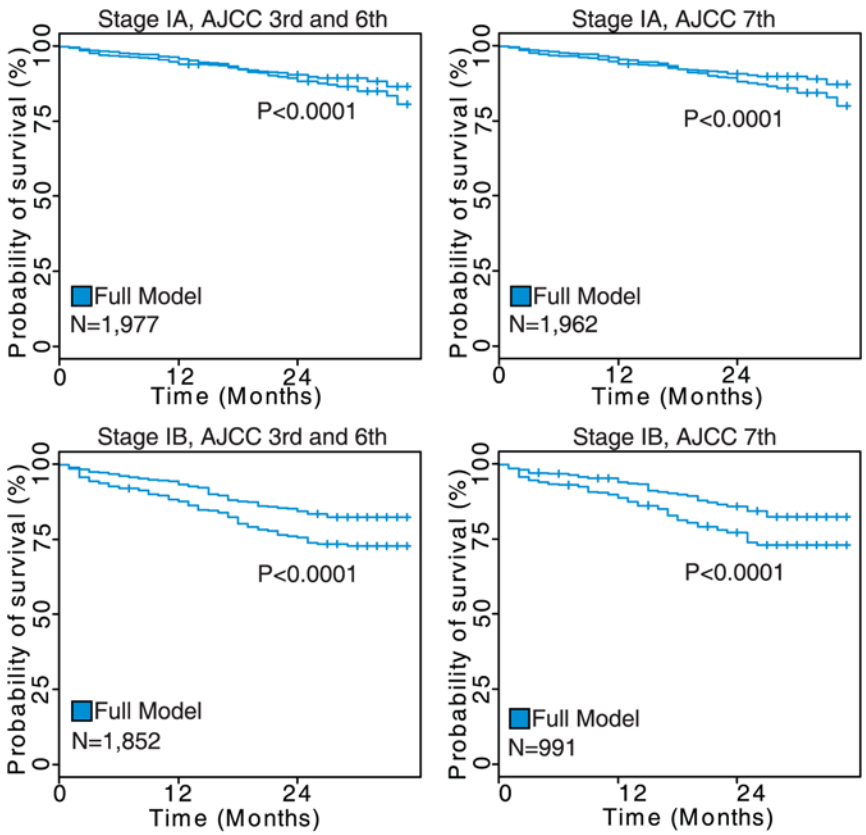

Figure 2.7. Results of survival analysis on squamous cell lung carcinoma patients diagnosed with Stage IA or IB disease. The Kaplan-Meier plots show the difference between low- and high-risk groups as determined by the comprehensive model. Data on sub-stage was only available for patients staged using the AJCC 6th Edition staging system (2004 and later) and for those patients converted into the 7th Edition.

External Validation

The comprehensive model was also able to improve prognostication in the external validation sets from MBRCC and Shedden et al (Shedden et al., 2008). Patients with both adenocarcinomas (n=794) and squamous cell carcinomas (n=758) were available from the MBRCC cohort, with only adenocarcinoma patients (n=439) available from the Director's Challenge cohort (Shedden et al., 2008). The comprehensive model performed consistently better across all three metrics considered when the models estimated using the SEER cohort was applied to the cohorts from MBRCC and the Director's Challenge study, with the results being consistent across histology in the MBRCC cohort. The model appeared to perform much better in the MBRCC cohort, possibly due to the lack of stage IV patients in the Director's Challenge cohort. These results are summarized in Tables 2.3, 2.4, and 2.5. Improvement in the construction of risk groups when using the comprehensive model is shown in Figure 2.8

Table 2.3. Harrell's C-statistics from each model for each of the patient cohorts, separated into AJCC coding system, treatment modality, and histology where possible.

	Adenocarcinoma				Squamous Cell Carcinoma			
	AJCC 3+6		*AJCC 7*		*AJCC 3+6*		*AJCC 7*	
	FM	SO	FM	SO	FM	SO	FM	SO
All SEER Patients	0.732	0.694	0.763	0.731	0.722	0.706	0.733	0.717
Surgery Only	0.768	0.723	0.742	0.707	0.762	0.742	0.689	0.670
Radiation Only	0.631	0.608	0.665	0.632	0.647	0.636	0.666	0.658
Both Surg+Rad	0.688	0.677	0.696	0.678	0.688	0.674	0.682	0.663
No Treatment	0.601	0.542	0.607	0.558	0.582	0.567	0.598	0.580
MBRCC	0.721	0.708	N/A	N/A	0.695	0.681	N/A	N/A
Dir's Challenge	0.687	0.660	N/A	N/A	N/A	N/A	N/A	N/A

Table 2.4. Nagelkerke's R^2 values from each model for each of the patient cohorts, separated into AJCC coding system, treatment modality, and histology where possible.

	Adenocarcinoma				**Squamous Cell Carcinoma**			
	AJCC 3+6		*AJCC 7*		*AJCC 3+6*		*AJCC 7*	
	FM	SO	FM	SO	FM	SO	FM	SO
All SEER Patients	0.294	0.253	0.305	0.274	0.289	0.274	0.246	0.230
Surgery Only	0.225	0.173	0.094	0.073	0.283	0.268	0.064	0.055
Radiation Only	0.107	0.084	0.140	0.115	0.109	0.103	0.120	0.118
Both Surg+Rad	0.204	0.178	0.084	0.072	0.201	0.184	0.095	0.089
No Treatment	0.066	0.034	0.075	0.044	0.051	0.042	0.065	0.054
MBRCC	0.343	0.311	N/A	N/A	0.244	0.233	N/A	N/A
Dir's Challenge	0.189	0.162	N/A	N/A	N/A	N/A	N/A	N/A

Table 2.5. Brier Scores from each model for each of the patient cohorts, separated into AJCC coding system, treatment modality, and histology where possible.

	Adenocarcinoma				**Squamous Cell Carcinoma**			
	AJCC 3+6		*AJCC 7*		*AJCC 3+6*		*AJCC 7*	
	FM	SO	FM	SO	FM	SO	FM	SO
All SEER Patients	0.134	0.143	0.144	0.150	0.119	0.119	0.162	0.161
Surgery Only	0.206	0.210	0.089	0.092	0.186	0.188	0.113	0.114
Radiation Only	0.097	0.099	0.163	0.168	0.102	0.102	0.170	0.172
Both Surg+Rad	0.096	0.097	0.153	0.154	0.098	0.099	0.160	0.163
No Treatment	0.098	0.101	0.167	0.178	0.078	0.081	0.151	0.152
MBRCC	0.071	0.074	N/A	N/A	0.079	0.081	N/A	N/A
Dir's Challenge	0.163	0.170	N/A	N/A	N/A	N/A	N/A	N/A

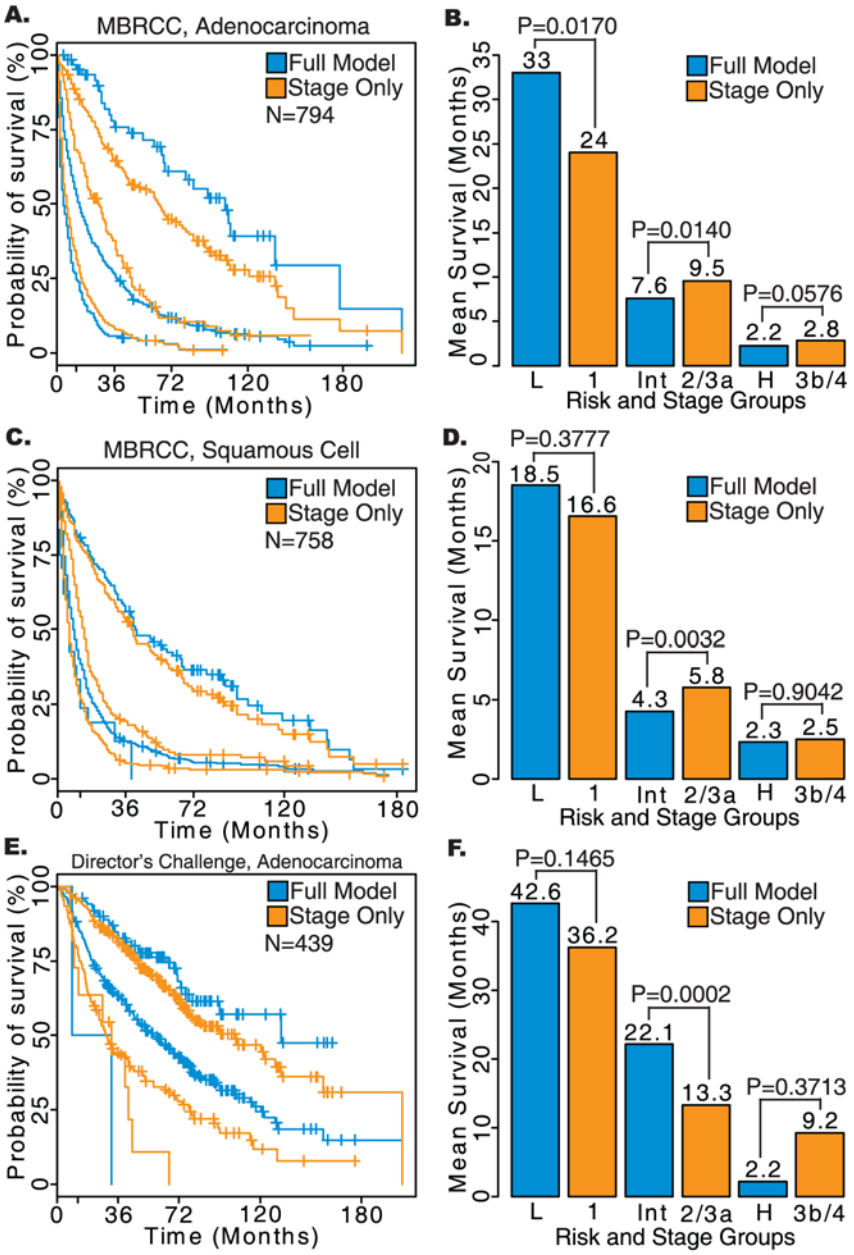

Figure 2.8. Results of survival analyses performed on patient cohorts from the Director's Challenge Study and the Mary Babb Randolph Cancer Center at West Virginia University.

Discussion

Substantial efforts have been made to establish prognostic factors for patients with lung cancer during the last two decades. The traditional prognostic factors are tumor size, vascular invasion, poor differentiation, high tumor-proliferative index, and genetic alterations, including K-*ras* (Rodenhuis et al., 1987; Slebos et al., 1990) and p53 (Horio et al., 1993). With the development of molecular biotechnology, especially high-throughput microarrays, there have been a number of promising studies on lung cancer prognosis by transcriptional profiling (Beer et al., 2002; Bhattacharjee et al., 2001; Bild et al., 2006; Borczuk et al., 2005; Chen et al., 2007; Raponi et al., 2006; Shedden et al., 2008; Guo et al., 2006; Guo et al., 2008). Although the traditional prognostic factors lack the information about the biological diversity of lung cancer and have not reflected the complexity of molecular mechanisms of these diseases, they are still the most valuable criteria for clinicians to decide the relevant therapies (Ludwig & Weinstein, 2005). For instance, Adjuvant! (www.adjuvantonline.com) is a prognostic system for lung cancer, breast cancer, and colon cancer based on traditional pathological features, including age, tumor stage, and grade. It has been independently validated as a reliable aid to clinical decision-making on average breast cancer patients (Olivotto et al., 2005). A study by Birim and others (Birim et al., 2006) also demonstrated that clinical factors outside of stage such as respiratory function, comorbidity, and smoking behaviors could be used to refine prognosis in a smaller cohort (N=766) of NSCLC patients.

In this study, we sought to investigate the impact of clinical, pathological, and demographic on lung cancer survival using a population-based approach. It was found that the addition of pathological and demographic variables to AJCC staging was able to significantly improve predictive ability in both lung adenocarcinomas and squamous cell carcinomas. These additional variables accounted for previously unexplained variation within and independent of stage, and resulted in a more accurate assessment of the risk for tumor recurrence when evaluated as integrated prognostic indicators. This effect persisted within multiple treatment modalities.

The comprehensive model was also able to improve prediction in stage I surgical adenocarcinoma patients, with the low-risk group selected by the comprehensive model having significantly better survival than the same group chosen using stage alone (P=0.0008). These patients may not benefit from additional therapies while, conversely, those who failed to be selected as low-risk may benefit from adjuvant chemotherapy.

These results demonstrate that the comprehensive prognostic model was able to reliably identify stage I NSCLC patients at higher risk for tumor recurrence, possibly benefitting from adjuvant chemotherapy. The external validation results indicate that the comprehensive prognostic model constructed from SEER population data could improve prognosis in multiple local hospitals.

The improvement seen is likely an underestimate, as the coefficients for stage only model, while intended to be representative of current clinical practice, were estimated with the same rigor as the comprehensive model likely resulting in a far more refined estimate of survival than could be garnered from clinical experience. Furthermore, there was a similar improvement in prediction when cases were converted into the AJCC 7th staging edition despite smaller sample sizes and shorter follow-up, indicating that the improvement attributed to the additional variables is not merely an artifact of shortcomings in the AJCC staging system which were addressed in the previous round of revisions. These findings show promise for a clinical model capable of considering these impactful factors in a robust and intuitive manner.

A limitation of the study is that it assumes that the quality or effectiveness of the treatment patients receive is similar across the study sample. As the patients come from a large number of different institutions it is unlikely that the assumption holds true. A separate analysis utilizing median county income as a surrogate measure for socio-economic status showed that there was a significant positive association between median county income and survival independent of other clinical, demographic, or pathologic factors. However, this sort of measure is not appropriate for individual prediction and was omitted from the model.

This study demonstrated the feasibility of refining NSCLC prognosis using easily collectable patient information in a large retrospective population study. A significant limitation of the study was the lack of information on the use of chemotherapy and co-morbidities present at the time of diagnosis. It is expected that inclusion of data found in the linked SEER-Medicare database will more appropriately address these issues and allow for further refinement of the model. In future research, we plan to construct a comprehensive model to estimate treatment benefits of commonly used chemotherapy utilizing the SEER-Medicare data partition patients according to a specific treatment approach. We envision that this model could be combined with future clinically validated gene signatures for a more refined assessment of patient risk.

Chapter3:
Effect of COPD and Chemotherapy on an Integrative Prognostic Model

Joseph Putila, Nancy Lan Guo

Abstract

Background: Selection of an appropriate course of therapy is a key factor in prolonging survival in lung cancer, and relies heavily on an accurate assessment of a patient's risk of recurrence. This analysis extends the previously described prognostic model through the use of data on chemotherapy and co-morbidities, further refining prognostication.

Methods: Data on 34,203 adenocarcinoma and 26,967 squamous cell carcinoma patients were used to determine the contribution of Chronic Obstructive Pulmonary Disease (COPD) to prognostication in 30 treatment combinations. A Cox model including COPD was estimated on 1000 bootstrap samples, with the resulting model assessed on ROC, Brier Score, Harrell's C, and Nagelkerke's R^2 metrics in order to highlight improvements in prognostication over a previously defined model without COPD.

Results: The addition of COPD to the model was shown to improve prognostication in patient groups covering multiple treatment combinations including those treated with and without chemotherapy. For adenocarcinoma patients, there was an improvement seen in the overall patient sample and in patients without chemotherapy, including those receiving surgery only. For squamous cell carcinoma, an improvement was seen in both the overall patient sample and in patients receiving multiple types of chemotherapy. In many cases these improvements persisted in several of the treatment subsets on multiple measures. COPD was also able to stratify patients into significantly different high and low-risk groups independent of other factors in stage and treatment sub-sets.

Significance: Although COPD is known to influence the treatment received, the results suggest that those with the disorder are at higher risk when compared to patients receiving a similar treatment. This extension of the previously described prognostic model to include COPD status and chemotherapy was able to significantly improve prognostication.

Introduction

One of the most important factors influencing the survival of patients with lung and bronchus cancers is the selection of an appropriate course of therapy based on an accurate assessment of patient risk. However, the selection of a course of therapy is currently based largely on tumor stage alone despite the contribution of other factors to patient survival. In particular, comorbidities such as COPD can have a significant effect on long-term survival due to varying treatment candidacy, increased complication rate, or decreased treatment efficacy (Janssen-Heijnen et al., 1998; Janssen-Heijnen et al., 2004; Kiri et al., 2010), and the prevalence of this disease is elevated in lung cancer patients independent of smoking status (Young et al., 2009). A major impediment is determining the contribution of each of these factors in a comprehensive prognostic model, rather than as independent effects.

Comorbidities which affect lung function, such as COPD, are likely to influence post-operative survival independent of their effects on surgical candidacy due to the possibility of decreased lung function. The Forced Expiratory Volume in 1 second (FEV_1) of a patient, a potential indicator of COPD, was found to be a significant prognostic factor in a model controlling for other clinical variables such as nodal status (Dehing-Oberije et al., 2008; Dehing-Oberije et al., 2009). Although COPD and lung cancer are often seen together in patients due to their shared association with smoking behavior, neither condition is a clinical end-point for the other. As such, COPD remains viable as a potential prognostic factor despite any shared similarity with lung cancer in the etiology of the disease.

Previously, we described a prognostic model utilizing similar data derived from the SEER cancer registry initiative (Putila et al., 2011) designed to address the problem of developing an integrative model. The previous approach utilized clinical and demographic variables in a single model to achieve superior prognostication over a similar model using stage alone. The model was however limited by the data in that the use of specific chemotherapies and the presence of comorbidities could not be determined. This analysis extends the previous analysis by utilizing patient records from the linked SEER-Medicare database. These additional data allow for the role of COPD in prognostication to be determined across combinations of surgical, radiological, and chemotherapeutic treatments (Elixhauser et al., 1998; Warren et al., 2002b).

Materials and Methods

Patient data was obtained from the linked SEER-Medicare database, a combination of population-based registry data and billing histories for patients covered by Medicare. In short, data from participating Surveillance Epidemiology and End-Results (SEER) registries were linked with Medicare data through the use of social security number, census tract, age, and other identifying variables(Warren et al., 2002b). The resulting data contain information on treatments administered, co-morbidities present in the patient, clinical presentation, survival, and demographics. Criteria for inclusion in this analysis were a diagnosis of lung or bronchus cancer between 1991 and 2005, and complete information on age, race, gender, tumor grade and stage, as well as valid follow-up and billing history. This set of patients was further refined to only include those with tumors broadly classifiable as either squamous cell carcinoma or adenocarcinoma. Cases reported solely from autopsy or death certificate were excluded. In total, 34,203 patients with adenocarcinomas and 26,967 patients with squamous cell carcinomas met these criteria.

The administration of chemotherapy was determined using Healthcare Common Procedure Coding System (HCPCS) codes in combination with International Classification of Diseases (ICD) codes. For each patient, all records were searched for entries with an HCPCS code indicating that chemotherapy was administered. These entries were then cross referenced with the ICD primary and secondary diagnosis codes for that entry to ensure that the agent was being administered for the treatment of lung or bronchus cancers. Four specific agents, cisplatin, carboplatin, docetaxel, and paclitaxel were considered in addition to a broad group covering a variety of chemotherapeutic agents.

Patients were also analyzed according to variable administration of curative surgery and radiological treatment. Five groups were formed on the basis of surgical and radiological treatments; a group containing all patients regardless of treatment, a group for patients having surgery without radiation, another for those having radiation without surgery, a group with both surgery and radiation, a group of patients without any treatment listed. Surgical and radiological treatment group assignments were based on the presence of any curative treatment of that type in the SEER portion of the data.

Comorbidities were measured as components of a version of the Charlson Comorbidity Index (CCI) adapted for use with administrative data (Charlson et al., 1987; Deyo et al., 1992; Klabunde et al., 2000; Romano et al., 1993). Comorbidities were determined using code available from the NCI designed specifically for this SEER-Medicare dataset

(http://healthservices.cancer.gov/seermedicare/program/comorbidity.html), under the assumption that the presence of a treatment for a specific disease in the billing history was indicative of its presence in the patient at the time. All claims out to three years prior to the diagnosis with lung cancer were analyzed for specific sets of ICD-9 codes indicative of one of the conditions listed in the CCI. Upon finding a relevant ICD code for any of the 18 conditions included in the index, patients were flagged as being positive for the corresponding condition. The determination of chemotherapy use and co-morbidities was done using *SAS* version 9.2 in the PC environment.

Patient age, race, and sex were identified using information in the SEER portion of the data, with Black and Asian/Pacific Islander patients being compared to Whites as a reference group. Tumor grade was also identified from the SEER portion of the data. Tumor grades 1 and 2 were grouped together and served as a reference group for grades 3 and 4, which were also grouped in the Cox models. Tumor stage was listed in the SEER portion of the data in either the 3rd or 6th Edition of the AJCC staging system. Records coded using the 6th Edition were able to be recoded to the 7th Edition provided that they had complete and valid information on tumor size, extension, nodal involvement, and distant metastases. Tumor stage I was used as a reference group in the Cox models. The number and distribution of patient characteristics is detailed in Table 3.1.

Each co-morbid condition was assessed as an independent predictor of survival using Cox proportional hazards modeling in patients treated with surgery but without chemotherapy or radiation indicated in order to isolate the effects of comorbidity from those of disparate treatment benefit or treatment candidacy (Table 3.2). Additionally, the presence of COPD as determined via the analysis of administrative records was assessed as an independent predictor of survival by testing for significant stratification of Kaplan-Meier survival curves. Patients were split into two outcome groups based on COPD status, and separate survival curves were estimated and plotted (Figure 3.1). Again, only patients receiving surgery without radiation or chemotherapy were included in order to better isolate the effect of COPD from other effects resulting from disparate treatment candidacy. The significance of the difference in survival was determined using the G-rho family of tests, with a p-value less than 0.05 indicating a significant difference in the survival curves estimated for the two groups being compared. Patients were then further split into stage groups and the effect of COPD in these subsets was assessed for each group. A selection of these results can be seen in Figures 3.2 through 3.7.

Table 3.1. Distribution of demographic and clinical characteristics of patients diagnosed with adenocarcinoma or squamous cell carcinoma in the original and recoded AJCC staging systems.

Variable	Adenocarcinoma			Squamous Cell Carcinoma		
	AJCC 3rd	AJCC 6th	AJCC 7th	AJCC 3rd	AJCC 6th	AJCC 7th
Stage						
1	9,510 (35.6%)	3,127 (42%)	2,371 (33.3%)	7,110 (33%)	1,968 (36.5%)	1,466 (28.6%)
2	1,475 (5.5%)	580 (7.8%)	1,400 (19.7%)	1,272 (5.9%)	574 (10.6%)	958 (18.7%)
3a	2,610 (9.8%)	661 (8.9%)	1,145 (16.1%)	3,074 (14.2%)	650 (12%)	1,114 (21.7%)
3b	4,638 (17.3%)	919 (12.3%)	438 (6.2%)	4,928 (22.8%)	907 (16.8%)	496 (9.7%)
4	8,516 (31.8%)	2,167 (29.1%)	1,762 (24.8%)	5,188 (24.0%)	1296 (24%)	1,089 (21.3%)
Grade						
1	3,436 (12.8%)	1,189 (16%)	1,143 (16%)	1,000 (4.6%)	191 (3.5%)	180 (3.5%)
2	8,780 (32.8%)	2,892 (38.8%)	2,800 (39.3%)	8,384 (38.9%)	2,284 (42.3%)	2,182 (42.6%)
3	13,885 (51.9%)	3,235 (43.4%)	3,044 (42.8%)	11,713 (54.3%)	2,850 (52.8%)	2,694 (52.6%)
4	648 (2.4%)	138 (1.9%)	129 (1.8%)	475 (2.2%)	70 (1.3%)	67 (1.3%)
Age, Mean(SD)	73.4 (7.2)	74.0 (7.6)	73.9 (7.6)	73.5 (6.9)	74.1 (7.4)	74.2 (7.3)
Race						
White	23218 (86.8%)	6,515 (87.4%)	6,219 (87.4%)	18,456 (85.6%)	4,727 (87.6%)	4,490 (87.6%)
Black	1945 (7.3%)	531 (7.1%)	508 (7.1%)	2,232 (10.3%)	486 (9%)	459 (9.0%)
Asian/Pac. Islander	1586 (5.9%)	408 (5.5%)	389 (5.5%)	884 (4.1%)	182 (3.4%)	174 (3.4%)
Sex						
Male	13678 (51.1%)	3,556 (47.7%)	3,372 (47.4%)	14,111 (65.4%)	3,370 (62.5%)	3,186 (62.3%)
Female	13071 (48.9%)	3,898 (52.3%)	3,744 (52.6%)	7,461 (34.6%)	2,025 (37.5%)	1,937 (37.7%)
+COPD Diagnosis	7948 (29.7%)	2,647 (35.5%)	2,538 (35.7%)	8,521 (39.5%)	2,568 (47.6%)	2,431 (47.6%)

The distributions of COPD and other variables in the original model were assessed in patients with very long and very short survival, relative to other patients, to determine if certain characteristics were disparate between groups of patients with varied survival. This was accomplished by partitioning patients based on survival time and status, then using a t-test or test of proportions to compare the distributions of variables between each group. Again, only those patients who were treated with surgery without radiation or chemotherapy were included. Long survival was defined as greater than 60 months for the original 3rd Edition staging, and greater than 24 months for the 6th and recoded 7th Edition groups due to shortened follow-up. Short survival was defined as less than 24 months for the original 3rd Edition staging, and less than 12 months for the 6th and recoded 7th Edition groups (Table 3.3).

Cox modeling was then used to estimate a proportional hazards model for use in quantitatively assessing patient risk given multiple variables. Two separate models were estimated: a Full model containing information on patient age, race, sex, tumor grade and stage, and a COPD model containing the same variables with an additional indicator of the presence of COPD. For both the Full and COPD models, a total of one-thousand Cox proportional-hazard models were estimated using bootstrapped samples equal in size to the original patient cohort. The distributions of coefficients were assessed for normality, and the mean of each coefficient was taken. This set of coefficients formed the final model for each approach.

The models were then assessed on a total of four metrics; area under an ROC curve, Brier Score, Harrell's c, and Nagelkerke's R^2. The ROC measure represents the area under an integrated ROC curve out to 36 months for the original 3rd Edition AJCC staging, and 24 months for the 6th and recoded 7th Edition AJCC staging. The Harrell's c measure is similar to the ROC measure, but takes into account the aspect of time. Nagelkerke's R^2 is a generalized form of the coefficient of determination (R^2) suitable for survival models. The ROC, c-statistic, and R^2 measures all range from 0 to 1, with higher scores indicating better performance. The Brier score is a measure of the accuracy of survival predictions, and ranges from 0 to 1 with lower scores being better. Similar to the ROC measure, the Brier score was calculated at 36 months for the original AJCC staging and 24 months for the recoded AJCC staging. These estimations were performed on 100 bootstrapped patient cohorts with a two-tailed T test being used to assess significant differences between models on each test. All model estimations and assessments were performed using *R* version x64 2.15.0, with the *survival*, *risksetAUC*, *Design*, *rms*, *pec*, and *Hmisc* packages.

The final models to be used in the online tool were constructed using the coefficients estimated earlier for the models containing tumor stage, grade, age, race, sex, and COPD status. A total of six models, one for each AJCC Staging Edition and histology combination, were constructed. Cutoffs for each AJCC Edition and histology group were determined by selecting two cutoffs to partition patients into High and Low-Risk groups across the range of Hazard Scores in the total patient population and testing the difference in survival between the Full and Stage Only models in an iterative manner. The final cutoffs represent points at which the Full model shows the greatest improvement in selecting both High and Low-Risk patients in that group. This stratification is represented in Figure 3.8. For each new patient entered using the tool, the Hazard Score is estimated using the coefficients for each variable in the appropriate model. The final coefficients and graphical representation of long-term survival as determined by Hazard Score are shown for each model in Figures 3.9 through 3.14. A representative image of the web-based application of the final prognostic model can be seen in Figure 3.15.

Results

COPD showed significant prognostic ability on multiple measures, both as an independent predictor and in the presence of other predictors. Other co-morbid conditions also showed promise as independent predictors in a Cox model (Table 3.2). As an independent predictor, COPD status alone was able to significantly stratify patients into high and low-risk groups ($p<0.05$) in four of six groups (Figure 3.1), although small sample size in the newer squamous cell carcinoma groups may have impeded achieving a significant stratification. The stratification in squamous cell carcinoma cases coded in the original 6th Edition and those recoded to the 7th Edition of AJCC staging was not significant despite a small degree of separation, with COPD patients having slightly diminished survival concordant with the other significant groups. In the significant cases, those without COPD showed consistently and significantly better survival when compared to those with COPD across the entire length of available follow-up, indicating that the effects of COPD are manifested in both long and short-term survival (Figure 3.1).

Table 3.2. Result of modeling survival in a model with each comorbid condition as an independent predictor. Shown are conditions which confer significantly poorer survival in one or both of the histologies considered when the sample was restricted to patients receiving surgery without radiation or chemotherapy.

	Adenocarcinoma		Squamous Cell	
Condition	Odds Ratio	p-value	Odds Ratio	p-value
Congestive Heart Failure	1.37	<0.0001*	1.27	0.0003*
Peripheral Vascular Disease	1.22	0.0020*	1.03	0.7234
Cerebrovascular Disease	1.16	0.0152*	1.06	0.3874
COPD	1.24	<0.0001*	1.11	0.0140*
Diabetes with sequelae	1.26	0.0186*	1.06	0.5964
Chronic Renal Failure	1.18	0.4083	1.57	0.0002*
Cirrhosis	0.74	0.3000	1.95	0.0024*
Gastrointestinal Ulcers	1.33	0.0193*	1.22	0.1337

Table 3.3. Methodology for assigning patients to outcome groups based on survival time and status, for use in comparing the prevalence of COPD in the AJCC 3rd Edition staging scheme (A) and AJCC 6th Edition and recoded 7th Edition (B). Survival status is based on disease (lung and bronchus cancer) specific criteria.

A. Survival Time/Status

	Alive	*Deceased*
<24 Months	Intermediate	Short
>=60 Months	Long	Long

B. Survival Time/Status

	Alive	*Deceased*
<12 Months	Intermediate	Short
>=24 Months	Long	Long

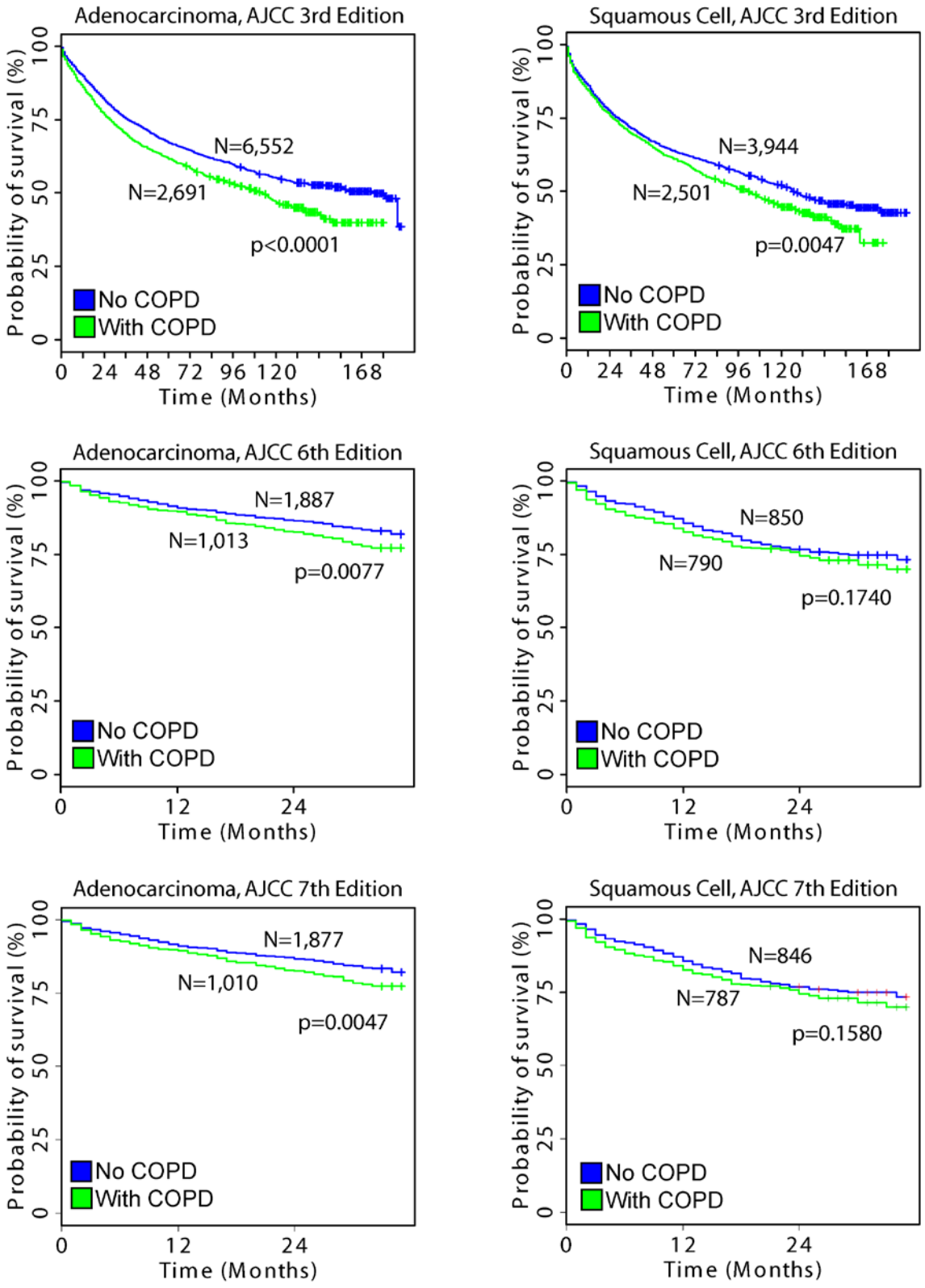

Figure 3.1. Survival curves generated for patients with and without COPD among those treated with surgery alone.

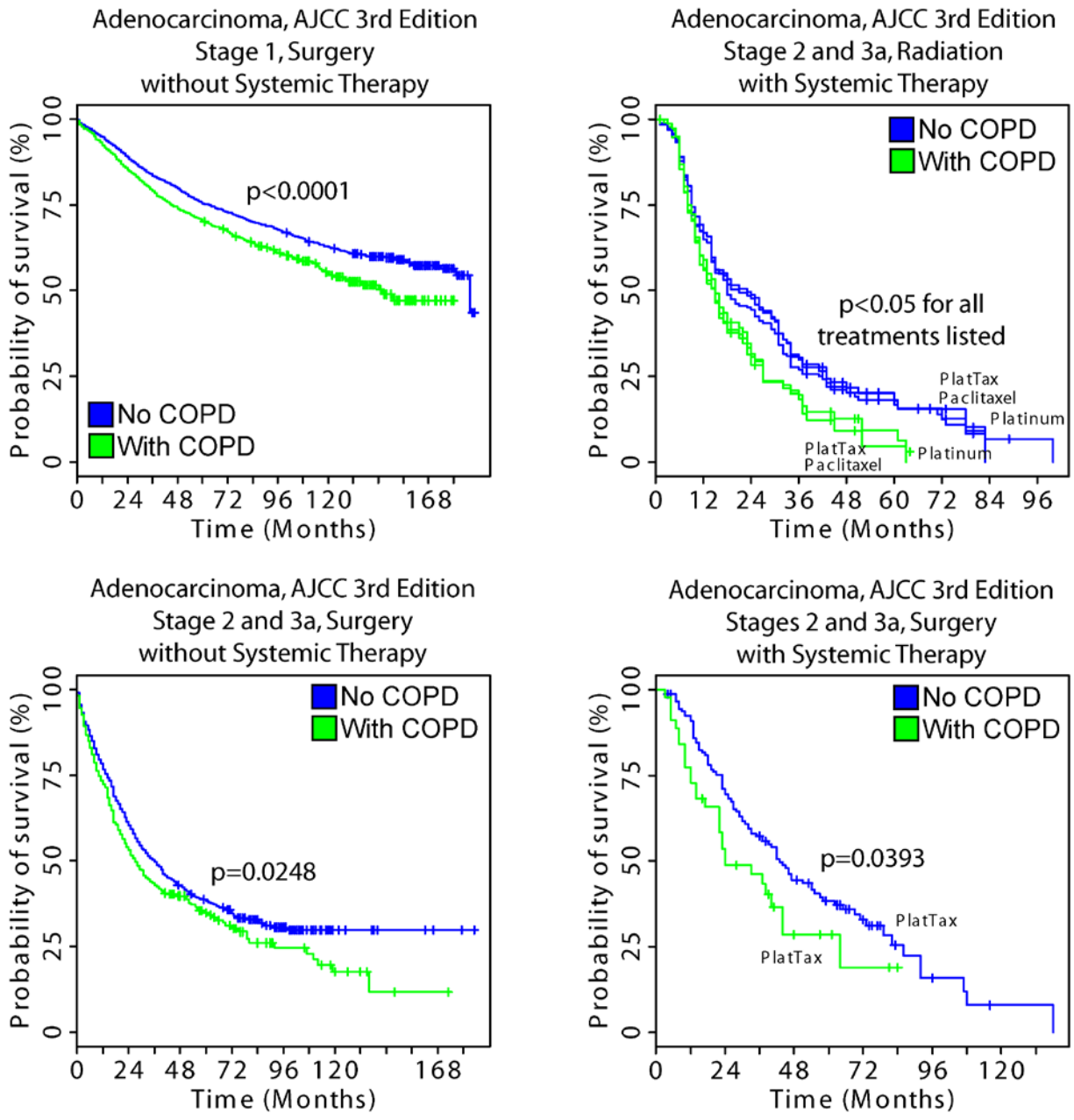

Figure 3.2. Effect of COPD in Adenocarcinoma AJCC 3rd Edition stage and treatment sub-groups. For each group, a clear and significant difference between the survival curves for patients with and without COPD can be seen, with patients identified as having COPD experiencing significantly poorer survival compared to those without COPD.

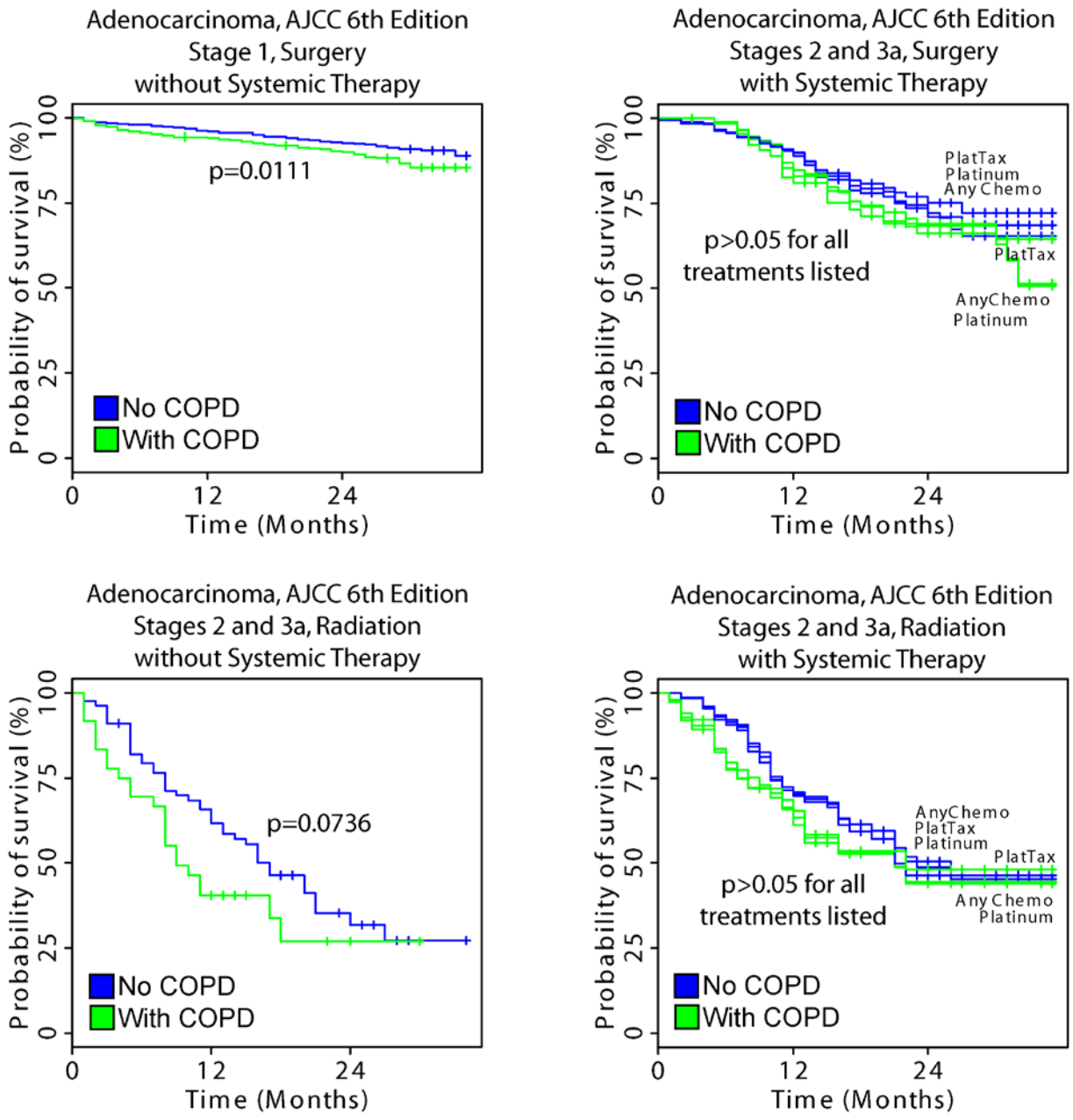

Figure 3.3. Effect of COPD in Adenocarcinoma AJCC 6th Edition stage and treatment sub-groups. For each group, patients without COPD tend to experience longer survival when compared to patients with COPD, although the difference is not significant in some cases shown.

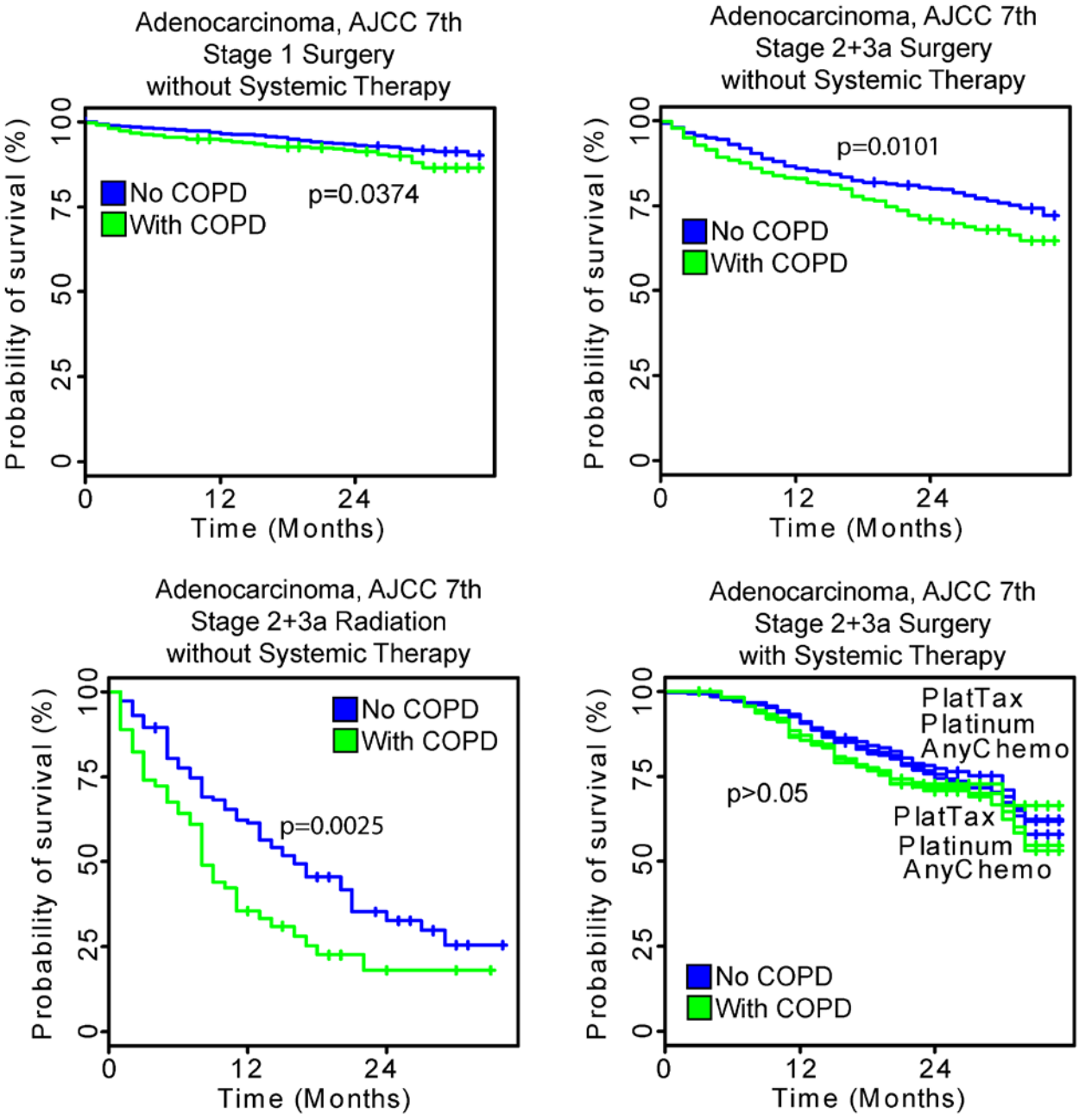

Figure 3.4. Effect of COPD in Adenocarcinoma AJCC 7th Edition stage and treatment sub-groups. For each group treated without chemotherapy, a clear and significant difference between the survival curves for patients with and without COPD can be seen, with patients identified as having COPD experiencing significantly poorer survival compared to those without the disease. The difference in survival for patients treated with systemic therapy was not significant, but trended toward patients with COPD having poorer survival.

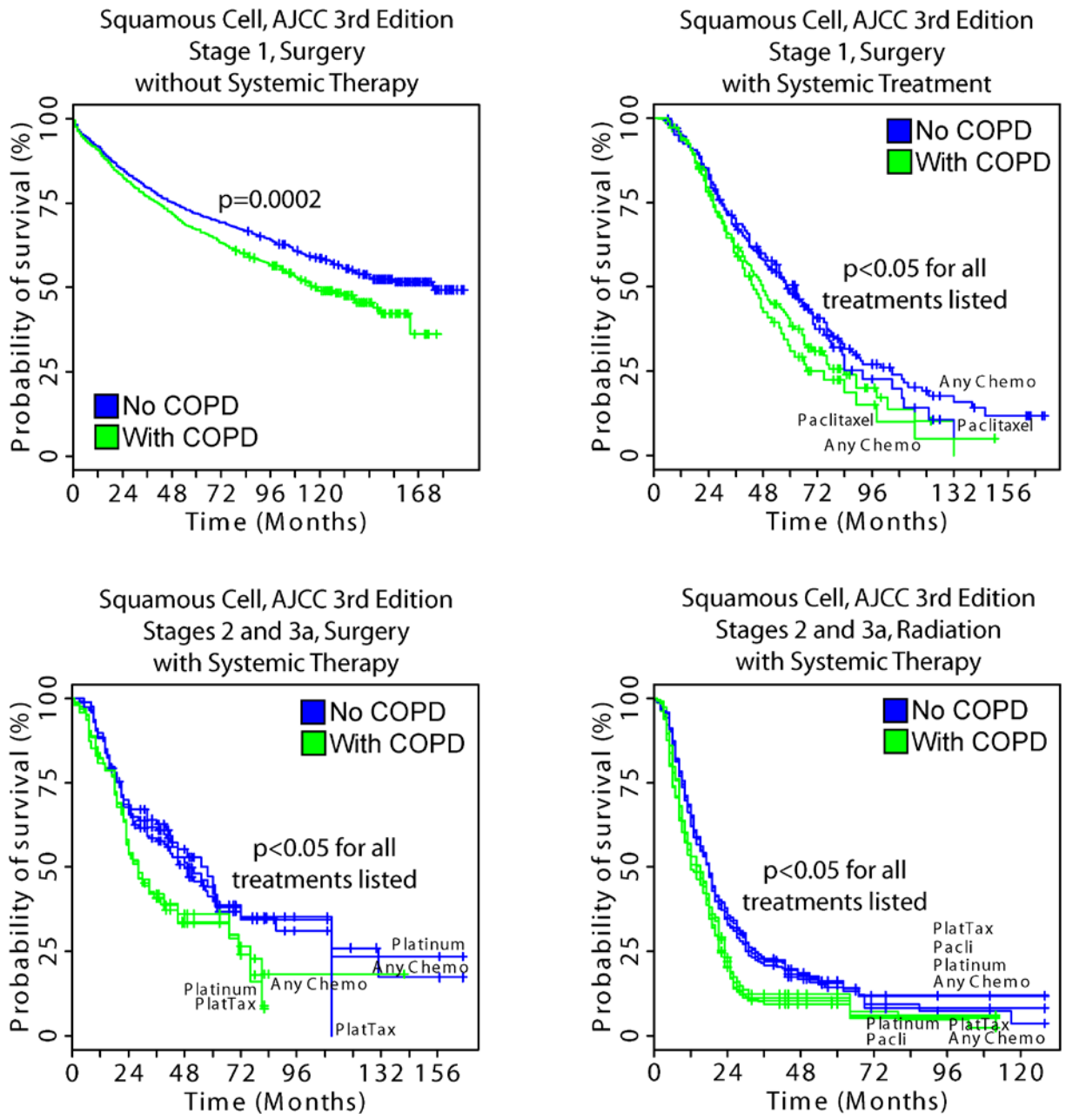

Figure 3.5. Effect of COPD in Squamous Cell AJCC 3rd Edition stage and treatment sub-groups. For each group, a clear and significant difference between the survival curves for patients with and without COPD can be seen, with patients identified as having COPD experiencing significantly poorer survival compared to those without the disease.

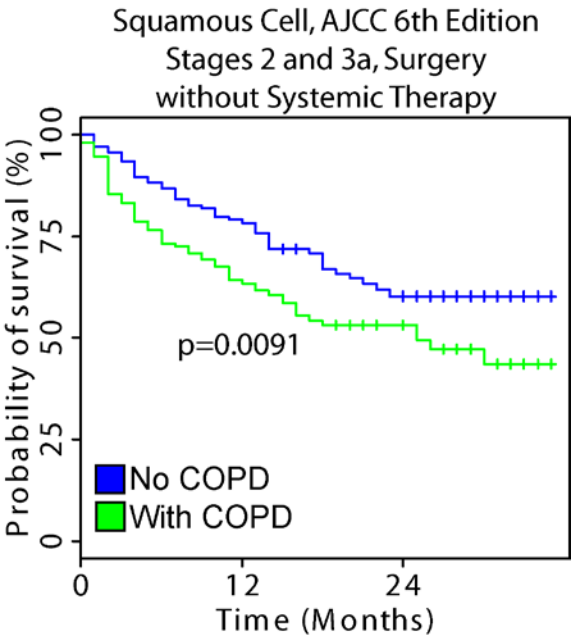

Figure 3.6. Effect of COPD in Squamous Cell AJCC 6th Edition stage and treatment sub-groups. For the group shown, a clear and significant difference between the survival curves for patients with and without COPD can be seen, with patients identified as having COPD experiencing significantly poorer survival compared to those without the disease.

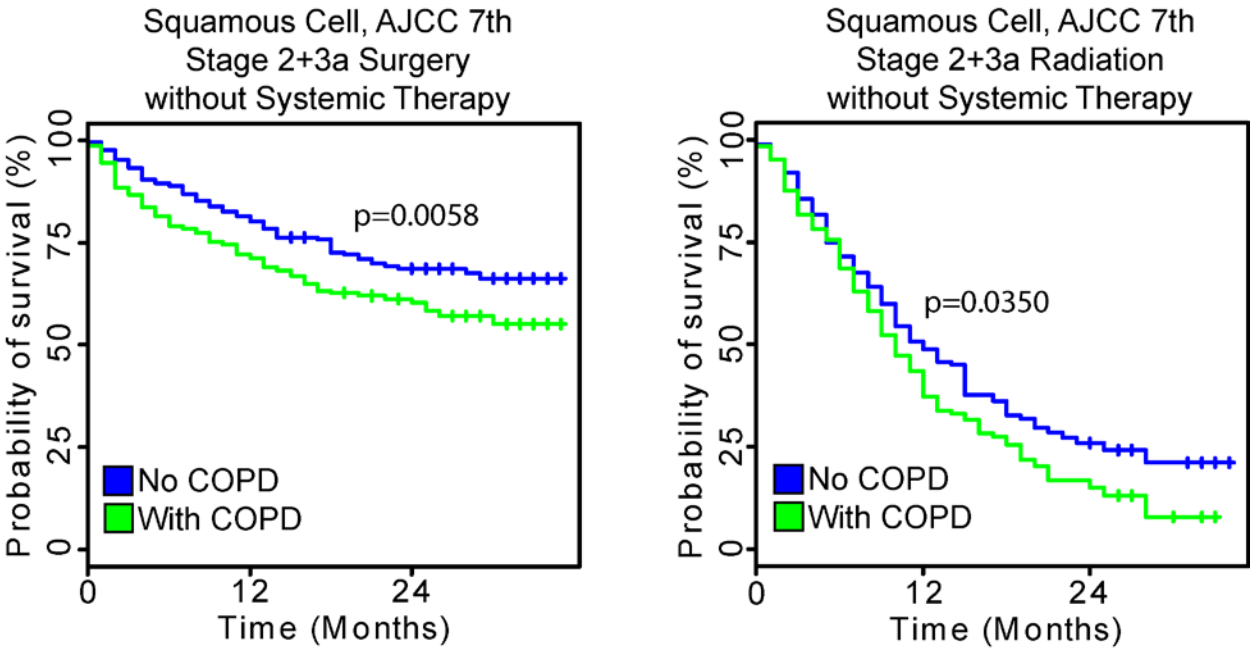

Figure 3.7. Effect of COPD in Squamous Cell AJCC 7th Edition stage and treatment sub-groups. For each group, a clear and significant difference between the survival curves for patients with and without COPD can be seen, with patients identified as having COPD experiencing significantly poorer survival compared to those without the disease.

The proportion of patients with COPD between those with relatively long and short survival was also assessed. Two survival cutoffs were used to split patients into three groups of short, intermediate, and long survival in order to test for differences in the distribution of each prognostic factor between groups of patients with relatively different survival. The difference in the prevalence of COPD between the short and long survival groups was assessed using a test of proportions. This test showed that COPD was much more prevalent in patients with relatively short survival when compared to those surviving relatively longer (Table 3.4). This was true for each histology and coding scheme, despite differences in cutoffs and length of follow-up between the original and recoded staging systems. When the same test was performed for the other covariates similar results were seen, with factors previously seen to favor increased or diminished survival being disparate between the groups. These results are summarized in Tables 3.4 through 3.9.

Patients were able to be split into high and low-risk groups with significantly different survival curves using COPD status alone in a variety of treatment and stage sub-groups. For adenocarcinoma patients staged using the original 3rd Edition staging, there was a significant difference in the survival of

Stage I patients treated with surgery alone. There was also a significant difference seen in Stage 2 and 3a patients treated with surgery with a platinum and taxane or without systemic therapy ($p<0.05$). A fourth stratification in Stage 2 and 3a patients treated with radiation and systemic therapy was also present ($p<0.05$, Figure 3.2).

In the group of patients staged using the original 6th Edition, Stage 2 and 3a radiation patients treated both with and without systemic therapy trended toward a similar stratification but did not achieve significance ($p>0.05$). A similar trend was observed for Stage 2 and 3a surgical patients treated with systemic therapy, although the degree of separation was again not significant ($p>0.05$). The group of stage 1 surgical patients treated without systemic therapy did however achieve a significant stratification ($p=0.0111$, Figure 3.3).

In the group of adenocarcinoma patients recoded to AJCC 7th Edition staging, there was a significant difference in survival between patients with and without COPD in Stage I surgical patients treated without systemic therapy ($p=0.0374$). This same difference was also present in Stage 2 and 3a surgical patients treated without systemic therapy ($p=0.0101$). There was a similar trend in the group of Stage 2 and 3a patients treated with surgery and systemic therapy, but the difference in survival was not significant for any of the systemic therapy groups ($p>0.05$). The group of Stage 2 and 3a patients without COPD also experienced significantly better survival when treated with radiation without systemic therapy ($p=0.0025$, Figure 3.4).

In squamous cell patients staged using the original 3rd Edition the difference in survival when stratifying on COPD status for Stage I surgical patients treated without systemic therapy was significant, with patients having COPD again experiencing poorer survival ($p=0.0002$). There was also a significant stratification in the corresponding group of Stage I surgical patients treated with systemic therapy ($p<0.05$). COPD Was able to produce a significant stratification in Stage 2 and 3a patients treated with systemic therapy in both the surgical and radiation groups ($p<0.05$, Figure 3.5). In patients staged using the original 6th Edition, COPD was able to produce a single stratification in Stage 2 and 3a surgical patients treated without systemic therapy ($p=0.0091$, Figure 3.6).

In the recoded 7th Edition staging, there were two groups where COPD was able to stratify patients. The first was in Stage 2 and 3a patients treated with surgery without any systemic therapy, with COPD patients having poorer survival ($p=0.0058$). The same pattern was observed in Stage 2 and 3a patients treated with radiation without any systemic therapy ($p=0.0350$, Figure 3.5).

Table 3.4. Proportion of patients with COPD in long and short-survival groups. A test of proportions was used to assess significant differences in the prevalence of COPD in the two survival groups created from patients treated with surgery without radiation or chemotherapy.

	Cutoffs		% COPD		
Group	Short	Long	Short	Long	P-Value
Adenocarcinoma 3rd	<24 mo	>= 60 mo	34.2%	24.2%	<0.0001
Adenocarcinoma 6th	<12 mo	>= 24 mo	39.0%	32.2%	0.0438
Adenocarcinoma 7th	<12 mo	>= 24 mo	39.5%	32.2%	0.0326
Squamous Cell 3rd	<24 mo	>= 60 mo	39.9%	34.0%	0.0005
Squamous Cell 6th	<12 mo	>= 24 mo	53.5%	44.9%	0.0368
Squamous Cell 7th	<12 mo	>= 24 mo	53.6%	44.8%	0.0344

Table 3.5. Mean AJCC tumor stage in long and short-survival groups. A t-test was used to assess significant differences in mean stage in the two survival groups created from patients treated with surgery without radiation or chemotherapy.

	Cutoffs		Mean Stage		
Group	Short	Long	Short	Long	P-Value
Adenocarcinoma 3rd	<24 mo	>= 60 mo	2.03	1.17	<0.0001
Adenocarcinoma 6th	<12 mo	>= 24 mo	2.22	1.29	<0.0001
Adenocarcinoma 7th	<12 mo	>= 24 mo	2.26	1.42	<0.0001
Squamous Cell 3rd	<24 mo	>= 60 mo	1.91	1.26	<0.0001
Squamous Cell 6th	<12 mo	>= 24 mo	1.97	1.28	<0.0001
Squamous Cell 7th	<12 mo	>= 24 mo	2.12	1.42	<0.0001

Table 3.6. Mean tumor grade in long and short-survival groups. A t-test was used to assess significant differences in mean tumor grade in the two survival groups created from patients treated with surgery without radiation or chemotherapy.

	Cutoffs		Mean Grade		
Group	Short	Long	Short	Long	P-Value
Adenocarcinoma 3rd	<24 mo	>= 60 mo	2.46	2.08	<0.0001
Adenocarcinoma 6th	<12 mo	>= 24 mo	2.38	2.03	<0.0001
Adenocarcinoma 7th	<12 mo	>= 24 mo	2.40	2.03	<0.0001
Squamous Cell 3rd	<24 mo	>= 60 mo	2.53	2.50	0.1486
Squamous Cell 6th	<12 mo	>= 24 mo	2.49	2.42	0.1374
Squamous Cell 7th	<12 mo	>= 24 mo	2.42	2.50	0.1201

Table 3.7. Mean patient age in long and short-survival groups. A t-test was used to assess significant differences in patient age in the two survival groups created from patients treated with surgery without radiation or chemotherapy.

	Cutoffs		Mean Age		
Group	Short	Long	Short	Long	P-Value
Adenocarcinoma 3rd	<24 mo	>= 60 mo	73.80	72.30	<0.0001
Adenocarcinoma 6th	<12 mo	>= 24 mo	75.03	73.40	0.0029
Adenocarcinoma 7th	<12 mo	>= 24 mo	74.85	73.42	0.0084
Squamous Cell 3rd	<24 mo	>= 60 mo	73.85	72.13	<0.0001
Squamous Cell 6th	<12 mo	>= 24 mo	74.80	73.34	0.001
Squamous Cell 7th	<12 mo	>= 24 mo	74.82	73.35	0.0101

Table 3.8. Proportion of patients classified as API (top) or Black (bottom) in long and short-survival groups. A test of proportions was used to assess significant differences in the prevalence of minority groups in the two survival groups created from patients treated with surgery without radiation or chemotherapy.

	Cutoffs		Race (API)		
Group	Short	Long	Short	Long	P-Value
Adenocarcinoma 3rd	<24 mo	>= 60 mo	4.70%	5.70%	0.1445
Adenocarcinoma 6th	<12 mo	>= 24 mo	6.18%	6.58%	0.9219
Adenocarcinoma 7th	<12 mo	>= 24 mo	6.32%	6.50%	1
Squamous Cell 3rd	<24 mo	>= 60 mo	3.54%	3.51%	1
Squamous Cell 6th	<12 mo	>= 24 mo	3.98%	2.95%	0.6176
Squamous Cell 7th	<12 mo	>= 24 mo	4.00%	3.00%	0.606

	Cutoffs		Race (Black)		
Group	Short	Long	Short	Long	P-Value
Adenocarcinoma 3rd	<24 mo	>= 60 mo	6.40%	5.10%	0.0865
Adenocarcinoma 6th	<12 mo	>= 24 mo	6.18%	4.96%	0.5221
Adenocarcinoma 7th	<12 mo	>= 24 mo	6.32%	4.97%	0.4728
Squamous Cell 3rd	<24 mo	>= 60 mo	8.03%	8.68%	0.5516
Squamous Cell 6th	<12 mo	>= 24 mo	4.87%	6.69%	0.4324
Squamous Cell 7th	<12 mo	>= 24 mo	4.91%	6.71%	0.4448

Table 3.9. Proportion of male patients in the long and short-survival groups. A test of proportions was used to assess significant differences in sex in the two survival groups created from patients treated with surgery without radiation or chemotherapy.

	Cutoffs		**Sex (Male)**		
Group	Short	Long	Short	Long	P-Value
Adenocarcinoma 3rd	<24 mo	>= 60 mo	56.50%	39.90%	<0.0001
Adenocarcinoma 6th	<12 mo	>= 24 mo	52.12%	40.94%	0.0014
Adenocarcinoma 7th	<12 mo	>= 24 mo	52.17%	40.92%	0.0014
Squamous Cell 3rd	<24 mo	>= 60 mo	68.31%	59.08%	<0.0001
Squamous Cell 6th	<12 mo	>= 24 mo	64.16%	55.71%	0.03916
Squamous Cell 7th	<12 mo	>= 24 mo	64.29%	55.62%	0.0350

Model Improvement in Adenocarcinoma

When using the original 3rd Edition staging system the addition of COPD to the model resulted in a significant improvement on the Harrell's c-statistic in the total group of patients treated with any treatment and those treated with surgery alone ($p<0.05$). This improvement on Harrell's c persisted for those groups when only patients treated without indication of chemotherapy were considered. There were also improvements with the addition of COPD on the ROC measure in patients treated without indication of chemotherapy ($p<0.05$). There were no significant improvements in either the Nagelkerke's R^2 measure or the Brier score.

In the group of patients coded using the 6th Edition staging, there was a significant improvement in on the Harrell's c measure in the total group of patients ($p=0.0480$). There was an additional improvement in patients treated with surgery regardless of indication of chemotherapy on the ROC measure ($p=0.0475$). As was the case with the 3rd Edition staging system, there were no significant improvements in either the Nagelkerke's R^2 measure or the Brier Score.

When patients were recoded to the 7th Edition staging, there was a significant improvement on the ROC measure in the total group of patients ($p=0.0436$) which persisted when only patients receiving surgery regardless of indication of chemotherapy were considered ($p=0.0326$). These results strongly indicate that the greatest improvement in prognostication occurred in patients without indication of chemotherapy, with a concordant improvement in the entire patient sample. The results for the three staging systems are summarized in Tables 3.10 through 3.13. Patient survival and the final Cox model coefficients for each staging system can be seen in Figures 3.9 through 3.11.

Table 3.10. P-values estimated by comparing the Harrell's c-statistics from 100 bootstrapped samples using the Cox model generated on the entire patient sample for the original Comprehensive model, and the same model estimated with a COPD indicator added. Significant values are highlighted in yellow, with values showing degradation in prognostication with the addition of COPD bolded and italicized.

Adenocarcinoma 3rd

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0016	0.0002	0.7142	0.7462	0.7217	0.5776
Surgery Only	0.0043	0.0028	0.9519	0.7105	0.8789	0.8064
Radiation Only	0.3493	0.2363	0.7388	0.8425	0.2166	0.6873
Surg + Rad	0.8379	0.8601	0.7690	0.8968	0.9997	0.9028
No Treatment	0.3143	0.0702	0.4461	0.3056	0.7928	0.1980

Adenocarcinoma 6th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0480	0.0856	0.3955	0.5910	0.9437	0.7413
Surgery Only	0.3046	0.2051	0.8345	0.7036	0.9168	0.6176
Radiation Only	0.1499	0.6973	0.2356	0.7225	0.5726	0.6144
Surg + Rad	0.6971	0.7273	0.5079	0.4342	0.9827	0.6161
No Treatment	0.0700	0.1956	0.4042	0.7254	0.4114	0.6656

Adenocarcinoma 7th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0500	0.0711	0.3876	0.6259	0.7937	0.6621
Surgery Only	0.2257	0.1882	0.6936	0.6574	0.9999	0.7558
Radiation Only	0.4322	0.9241	0.8388	0.7593	0.4598	0.9601
Surg + Rad	0.6118	0.7944	0.6889	0.5453	0.7461	0.7422
No Treatment	0.0857	0.3129	0.3658	0.4242	0.6381	0.4773

Table 3.11. P-values estimated by comparing the Nagelkerke's R^2 statistic from 100 bootstrapped samples using the Cox model generated on the entire patient sample for the original Comprehensive model, and the same model estimated with a COPD indicator added. Significant values are highlighted in yellow, with values showing degradation in prognostication with the addition of COPD bolded and italicized.

Adenocarcinoma 3rd

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.1584	0.1201	0.9228	0.9007	0.9004	0.8727
Surgery Only	0.0866	0.1202	0.8685	0.9079	0.8471	0.8309
Radiation Only	0.5638	0.7560	0.4240	0.4644	0.2454	0.5246
Surg + Rad	0.5914	0.6995	0.8545	0.8178	0.9995	0.8949
No Treatment	0.4754	0.1228	0.5587	0.5011	0.5585	0.2461

Adenocarcinoma 6th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.2176	0.1477	0.6948	0.8449	0.9706	0.9137
Surgery Only	0.0874	0.2201	0.5911	0.5291	0.9314	0.6899
Radiation Only	0.1130	0.1663	0.8211	0.9646	0.5765	0.6126
Surg + Rad	0.4996	0.9854	0.2297	0.1906	0.9444	0.9301
No Treatment	0.3263	0.6722	0.7703	0.8269	0.4565	0.9344

Adenocarcinoma 7th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0814	0.3379	0.5324	0.6550	0.6325	0.5860
Surgery Only	0.0679	0.1815	0.5138	0.3998	0.9427	0.4235
Radiation Only	0.2211	0.3464	0.9340	0.7292	0.9713	0.3919
Surg + Rad	0.5132	0.9514	0.4026	0.4431	0.9588	0.8120
No Treatment	0.2814	0.5350	0.4753	0.6372	0.8534	0.6812

Table 3.12. P-values estimated by comparing the area under an Integrated ROC curve out to 36 months for the 3rd Edition and 24 months for the 6th and recoded 7th Edition from 100 bootstrapped samples using the Cox model generated on the entire patient sample for the original Comprehensive model, and the same model estimated with a COPD indicator added. Significant values are highlighted in yellow, with values showing degradation in prognostication with the addition of COPD bolded and italicized.

Adenocarcinoma 3rd

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.1453	0.0404	0.8635	0.7894	0.8856	0.7661
Surgery Only	0.0980	0.0269	0.9158	0.7657	0.7520	0.8878
Radiation Only	0.2674	0.3197	0.5190	0.5322	0.2366	0.4830
Surg + Rad	0.6427	0.6579	0.9475	0.9902	0.9664	0.9147
No Treatment	0.2912	0.0497	0.5786	0.4414	0.7798	0.3310

Adenocarcinoma 6th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0834	0.1222	0.5936	0.7844	0.9602	0.9752
Surgery Only	0.0475	0.1127	0.5290	0.5653	0.9689	0.6005
Radiation Only	0.0908	0.1167	0.6157	0.8712	0.8276	0.7574
Surg + Rad	0.4155	0.9416	0.3495	0.2381	0.9486	0.7450
No Treatment	0.2379	0.4202	0.7530	0.9481	0.6597	0.8425

Adenocarcinoma 7th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0436	0.1846	0.4478	0.7310	0.8017	0.7131
Surgery Only	0.0326	0.0630	0.3348	0.3295	0.7815	0.5313
Radiation Only	0.2037	0.4139	0.8399	0.9630	0.8433	0.6449
Surg + Rad	0.6475	0.7967	0.4103	0.3634	0.8988	0.8784
No Treatment	0.1568	0.2826	0.5004	0.7161	0.9540	0.5240

Table 3.13. P-values estimated by comparing the Brier score at 36 months for the 3rd Edition and 24 months for the 6th and recoded 7th Edition from 100 bootstrapped samples using the Cox model generated on the entire patient sample for the original Comprehensive model, and the same model estimated with a COPD indicator added. Significant values are highlighted in yellow, with values showing degradation in prognostication with the addition of COPD bolded and italicized.

Adenocarcinoma 3rd

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0801	0.1161	0.6341	0.7535	0.6077	0.6832
Surgery Only	0.3738	0.3481	0.7197	0.7455	0.6505	0.7429
Radiation Only	0.7452	0.7439	0.8391	0.8165	0.7534	0.8005
Surg + Rad	0.9379	0.8855	0.8526	0.9518	0.9409	0.8444
No Treatment	0.8303	0.7250	0.9825	0.9604	0.9445	0.8964

Adenocarcinoma 6th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.1731	0.1335	0.8574	0.7562	0.9032	0.6800
Surgery Only	0.6337	0.6007	0.8431	0.9083	0.9250	0.8183
Radiation Only	0.3224	0.2946	0.7990	0.8363	0.9905	0.7586
Surg + Rad	0.6245	0.6985	0.9679	0.9779	0.9738	0.7066
No Treatment	0.6505	0.6440	0.9686	0.7483	0.7406	0.9309

Adenocarcinoma 7th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.1273	0.0677	0.9072	0.5518	0.9375	0.5797
Surgery Only	0.5045	0.3737	0.8554	0.9413	0.9563	0.8841
Radiation Only	0.3143	0.2295	0.8211	0.9512	0.8165	0.6344
Surg + Rad	0.8202	0.8252	0.9340	0.9979	0.9321	0.7037
No Treatment	0.5261	0.4674	0.8933	0.7485	0.7839	0.9535

Model Improvement in Squamous Cell Carcinoma

Using the original AJCC 3rd Edition coding scheme, there were multiple treatment groups which showed a significant improvement in prognostication with the addition of COPD to the model. When any combination of surgical or radiological treatment was considered regardless of indication of chemotherapy, there was a significant improvement in the total sample on the Harrell's c-statistic (p=0.0130). This improvement persisted when only considering patients treated with any chemotherapy (p=0.0244) or with a platinum-based agent (p=0.0125, Table 3.14). A similar improvement was seen in patients treated with a platinum-based agent or with a platinum-based agent and a taxane and any other surgical or radiological treatment for Nagelkerke's R^2 and ROC ($p<0.05$, Tables 3.15 and 3.16). The addition of COPD to the model was able to significantly improve prognostication on Harrell's c-statistic in surgical patients across the range of chemotherapy sub-groups ($p<0.05$). There was also a significant improvement on the Brier score when all patients were considered (p=0.0239, Table 3.17).

In patients staged using the 6th Edition there was a significant degradation in model performance on Harrell's c-statistic in patients without an indication of treatment (p=0.0089). There were not any further differences in other measures for this staging system. When the patients were recoded using the AJCC 7th Edition criteria the difference in survival was again less pronounced than when using the original 3rd Edition staging criteria. There was a single significant improvement on the Brier score for patients treated without chemotherapy (p=0.0489). These results are summarized in Tables 3.14 through 3.17. The final model used and a representation of survival for squamous cell carcinoma patients can be seen in Figures 3.12 through 3.14.

Table 3.14. P-values estimated by comparing the Harrell's c-statistics from 100 bootstrapped samples using the Cox model generated on the entire patient sample for the original Comprehensive model, and the same model estimated with a COPD indicator added. Significant values are highlighted in yellow, with values showing degradation in prognostication with the addition of COPD bolded and italicized.

Squamous Cell 3rd

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0130	0.0884	0.0244	0.0125	0.1852	0.0571
Surgery Only	0.0006	0.0107	0.0058	0.0077	0.0333	0.0263
Radiation Only	0.4007	0.8252	0.1644	0.0867	0.2864	0.2079
Surg + Rad	0.3112	0.1602	0.7178	0.6892	0.5738	0.9322
No Treatment	0.3427	0.2802	0.4459	0.2366	0.6742	0.3606

Squamous Cell 6th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.6265	0.4355	0.6450	0.6121	0.9647	0.6468
Surgery Only	0.8196	0.9442	0.2354	0.1670	0.6093	0.1985
Radiation Only	0.8807	0.9912	0.8581	0.7801	0.7409	0.9108
Surg + Rad	0.8865	0.9543	0.6688	0.9064	0.9892	0.9165
No Treatment	***0.0089***	0.3086	0.4642	0.2662	0.6188	0.4358

Squamous Cell 7th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.3653	0.1168	0.7176	0.4349	0.9719	0.4815
Surgery Only	0.5885	0.3530	0.3557	0.1124	0.5362	0.1533
Radiation Only	0.9159	0.8404	0.9850	0.9178	0.6276	0.8035
Surg + Rad	0.9691	0.9854	0.6027	0.2893	0.1240	0.0891
No Treatment	0.0686	0.4795	0.4162	0.0971	0.3220	0.2295

Table 3.15. P-values estimated by comparing the Nagelkerke's R^2 statistic from 100 bootstrapped samples using the Cox model generated on the entire patient sample for the original Comprehensive model, and the same model estimated with a COPD indicator added. Significant values are highlighted in yellow, with values showing degradation in prognostication with the addition of COPD bolded and italicized.

Squamous Cell 3rd

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.1014	0.4553	0.1107	0.0229	0.1063	0.0286
Surgery Only	0.1290	0.2859	0.1533	0.0674	0.1132	0.0873
Radiation Only	0.6920	0.6309	0.0984	0.0986	0.2469	0.3527
Surg + Rad	0.5961	0.2976	0.8073	0.8086	0.2950	0.5920
No Treatment	0.7813	0.5202	0.5303	0.3370	0.9604	0.4286

Squamous Cell 6th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.2640	0.3448	0.8813	0.9451	0.7800	0.9195
Surgery Only	0.5186	0.3857	0.7070	0.6605	0.9333	0.6305
Radiation Only	0.8845	0.8187	0.6258	0.6304	0.9511	0.8436
Surg + Rad	0.6827	0.6423	0.7827	0.9323	0.9947	0.9845
No Treatment	0.5641	0.7819	0.8843	0.8359	0.8421	0.9833

Squamous Cell 7th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.2618	0.3960	0.9129	0.7735	0.7144	0.7504
Surgery Only	0.4343	0.2752	0.8333	0.5224	0.9445	0.5179
Radiation Only	0.9445	0.9359	0.8108	0.8063	0.7989	0.7206
Surg + Rad	0.4112	0.6121	0.6994	0.9492	0.7472	0.8845
No Treatment	0.8134	0.8734	0.6067	0.4735	0.7184	0.8235

Table 3.16. P-values estimated by comparing the area under an Integrated ROC curve out to 36 months for the 3rd Edition and 24 months for the 6th and recoded 7th Edition from 100 bootstrapped samples using the Cox model generated on the entire patient sample for the original Comprehensive model, and the same model estimated with a COPD indicator added. Significant values are highlighted in yellow, with values showing degradation in prognostication with the addition of COPD bolded and italicized.

Squamous Cell 3rd

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0823	0.3764	0.0694	0.0383	0.1061	0.0412
Surgery Only	0.0141	0.0718	0.0719	0.0553	0.0837	0.0811
Radiation Only	0.9420	0.5654	0.0445	0.0600	0.1807	0.1807
Surg + Rad	0.4149	0.2187	0.8758	0.9028	0.5063	0.7268
No Treatment	0.7607	0.5585	0.5663	0.3072	0.7893	0.5403

Squamous Cell 6th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.4187	0.3237	0.8530	0.8345	0.7866	0.9525
Surgery Only	0.3426	0.2310	0.7967	0.6588	0.9824	0.6829
Radiation Only	0.9241	0.7226	0.5262	0.5636	0.9425	0.8440
Surg + Rad	0.6037	0.6105	0.6826	0.9929	0.9136	0.9764
No Treatment	0.3139	0.6181	0.9072	0.7603	0.7930	0.8559

Squamous Cell 7th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.5050	0.3799	0.8761	0.7701	0.8408	0.8796
Surgery Only	0.2684	0.1335	0.8804	0.5706	0.8931	0.5738
Radiation Only	0.9690	0.8683	0.6810	0.7237	0.8795	0.9022
Surg + Rad	0.3648	0.5810	0.6488	0.9048	0.9345	0.9024
No Treatment	0.5305	0.9266	0.6311	0.4731	0.7956	0.6966

Table 3.17. P-values estimated by comparing the Brier score at 36 months for the 3rd Edition and 24 months for the 6th and recoded 7th Edition from 100 bootstrapped samples using the Cox model generated on the entire patient sample for the original Comprehensive model, and the same model estimated with a COPD indicator added. Significant values are highlighted in yellow, with values showing degradation in prognostication with the addition of COPD bolded and italicized.

Squamous Cell 3rd

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0239	0.1278	0.2410	0.1933	0.3010	0.2519
Surgery Only	0.4963	0.6791	0.4563	0.3227	0.5379	0.3675
Radiation Only	0.8111	0.8972	0.7258	0.7170	0.7838	0.8460
Surg + Rad	0.2956	0.1504	0.9136	0.7881	0.6282	0.6697
No Treatment	0.9588	0.9889	0.9258	0.9154	0.9958	0.9171

Squamous Cell 6th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.1250	0.0973	0.8682	0.9657	0.9154	0.8012
Surgery Only	0.6225	0.5191	0.9971	0.9447	0.9593	0.9295
Radiation Only	0.7891	0.9429	0.9727	0.9776	0.9543	0.8708
Surg + Rad	0.5036	0.4595	0.6109	0.8677	0.8574	0.8006
No Treatment	0.9887	0.8279	0.8543	0.8139	0.8145	0.8024

Squamous Cell 7th

	Total	No Chemo	Any Chemo	Platinum	Paclitaxel	Plat+Tax
Any Treatment	0.0958	0.0489	0.8356	0.9966	0.8708	0.8285
Surgery Only	0.5874	0.4013	0.9734	0.8271	0.8533	0.8546
Radiation Only	0.7728	0.8855	0.9251	0.9958	0.9410	0.9411
Surg + Rad	0.3010	0.4205	0.6278	0.8229	0.7921	0.7954
No Treatment	0.9410	0.7603	0.8705	0.7998	0.8941	0.7836

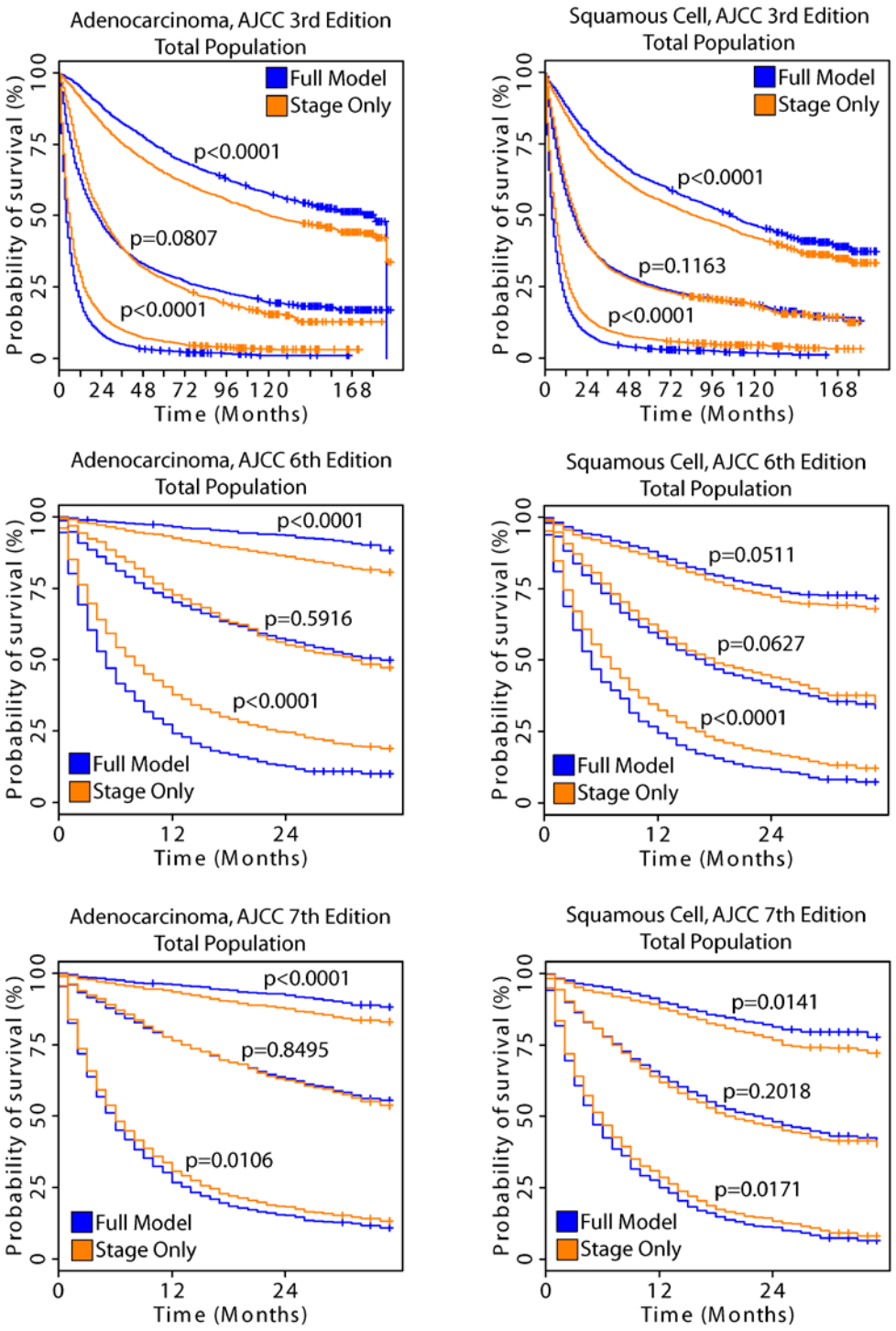

Figure 3.8. Improvement in the Full model using COPD over Stage Alone. For each Kaplan-Meier plot, the three pairs of lines represent the High, Intermediate, and Low-Risk groups defined for each of the two models shown. The model using only AJCC Stage is shown in orange, while the Full model with COPD status added is shown in blue. For each plot shown, the Full model with COPD status was able to produce a Low-Risk group with better survival and a High-Risk group with poorer survival, with most cases being significant (p<0.05).

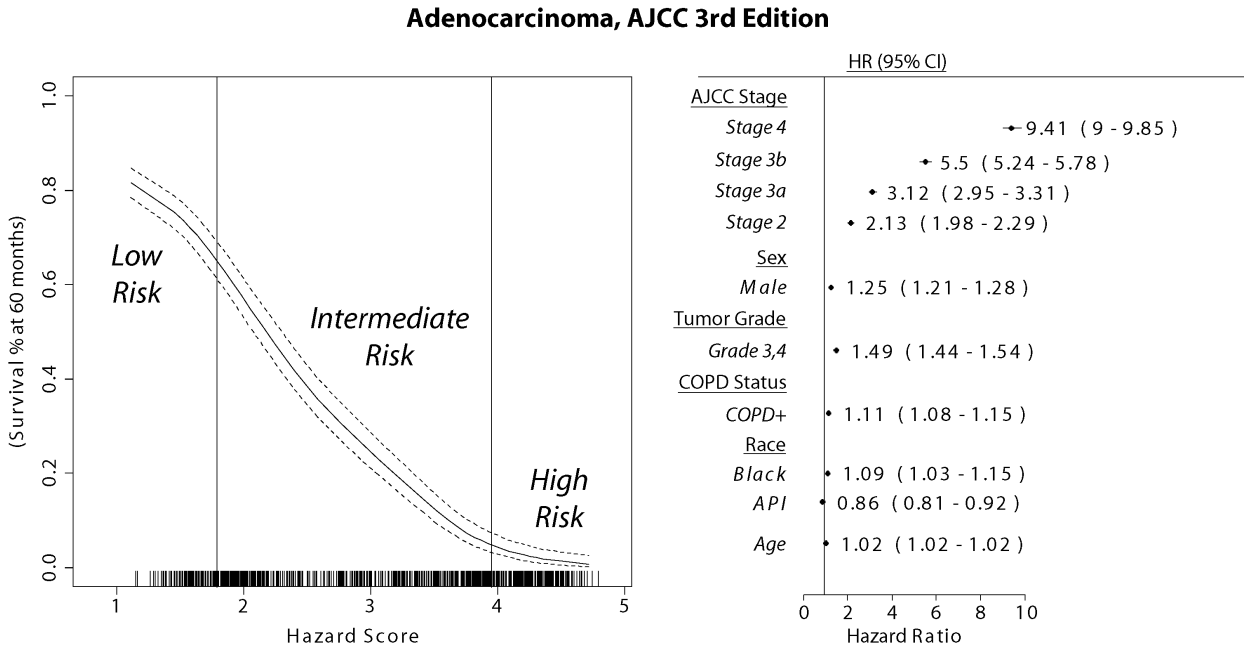

Figure 3.9. Final model for Adenocarcinoma AJCC 3rd Edition. Model coefficients used to determine the Hazard Score for each patient are shown on the forest plot (right). Patient survival at 60 months for the total population sample is shown for the range of Hazard Scores (left), with the risk-groups delimited by vertical bars.

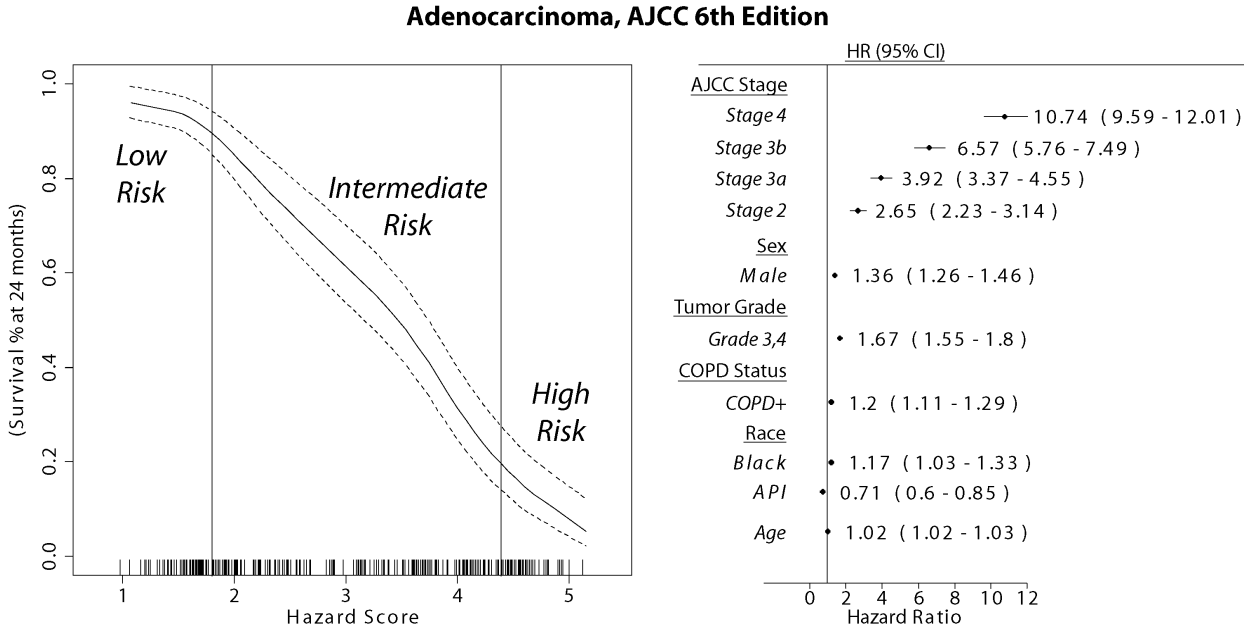

Figure 3.10. Final model for Adenocarcinoma AJCC 6th Edition. Model coefficients used to determine the Hazard Score for each patient are shown on the forest plot (right). Patient survival at 24 months for the total population sample is shown for the range of Hazard Scores (left), with the risk-groups delimited by vertical bars.

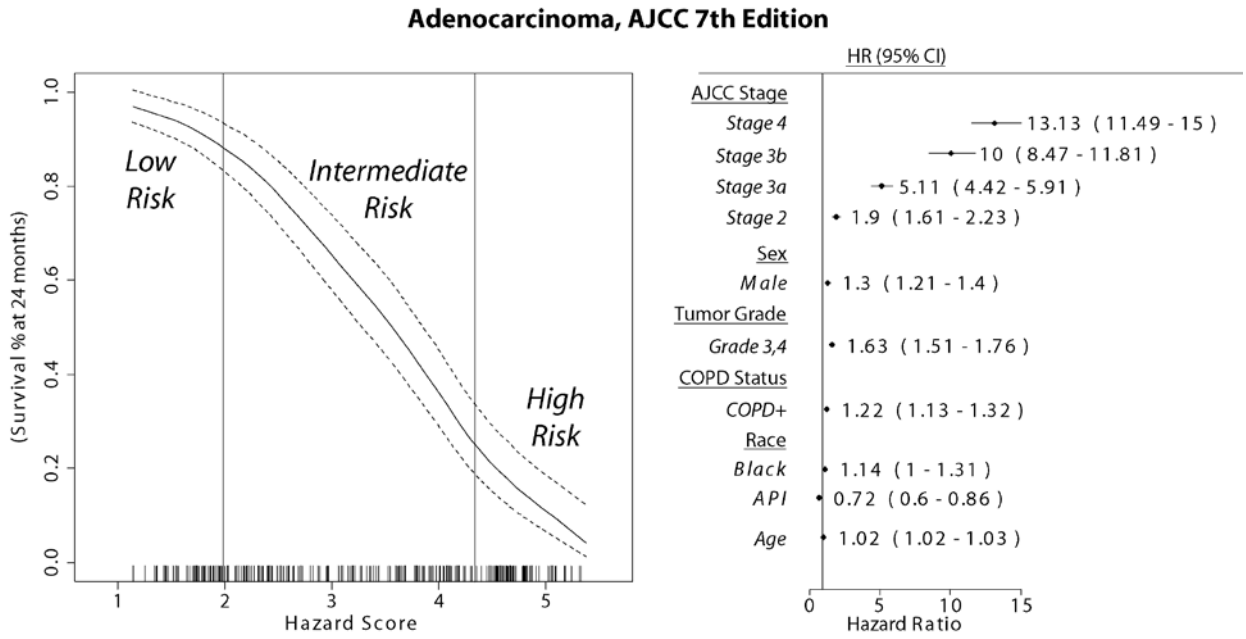

Figure 3.11. Final model for Adenocarcinoma AJCC 7th Edition. Model coefficients used to determine the Hazard Score for each patient are shown on the forest plot (right). Patient survival at 24 months for the total population sample is shown for the range of Hazard Scores (left), with the risk-groups delimited by vertical bars.

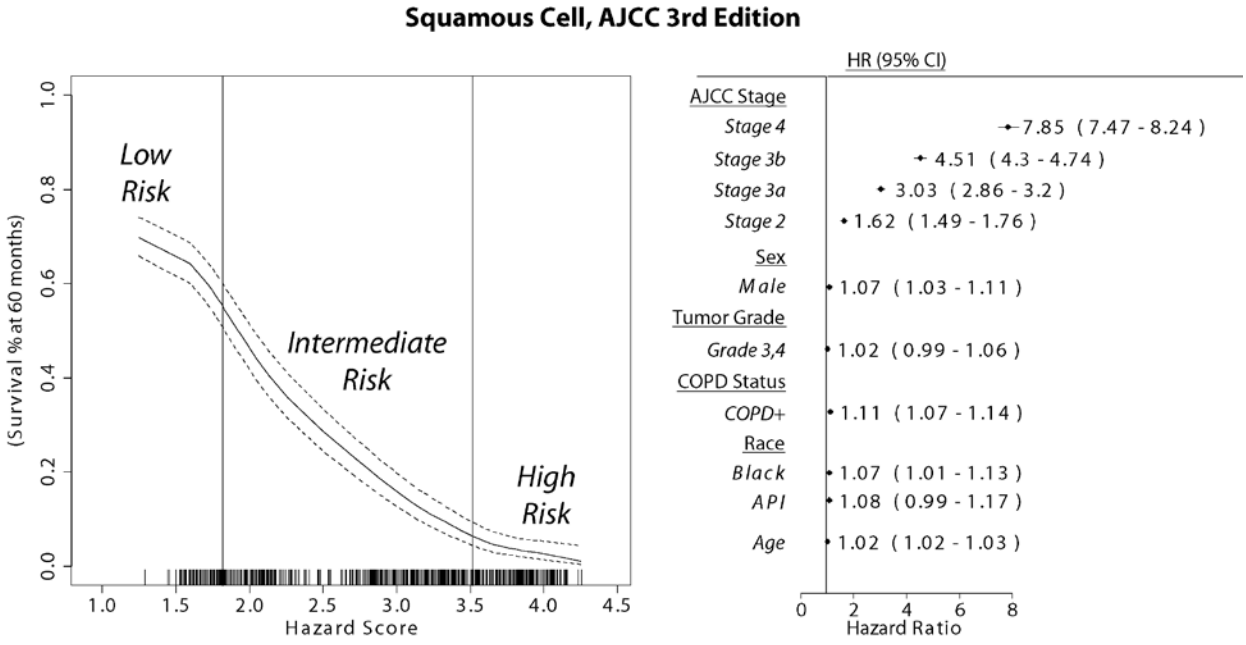

Figure 3.12. Final model for Squamous Cell AJCC 3rd Edition. Model coefficients used to determine the Hazard Score for each patient are shown on the forest plot (right). Patient survival at 60 months for the total population sample is shown for the range of Hazard Scores (left), with the risk-groups delimited by vertical bars.

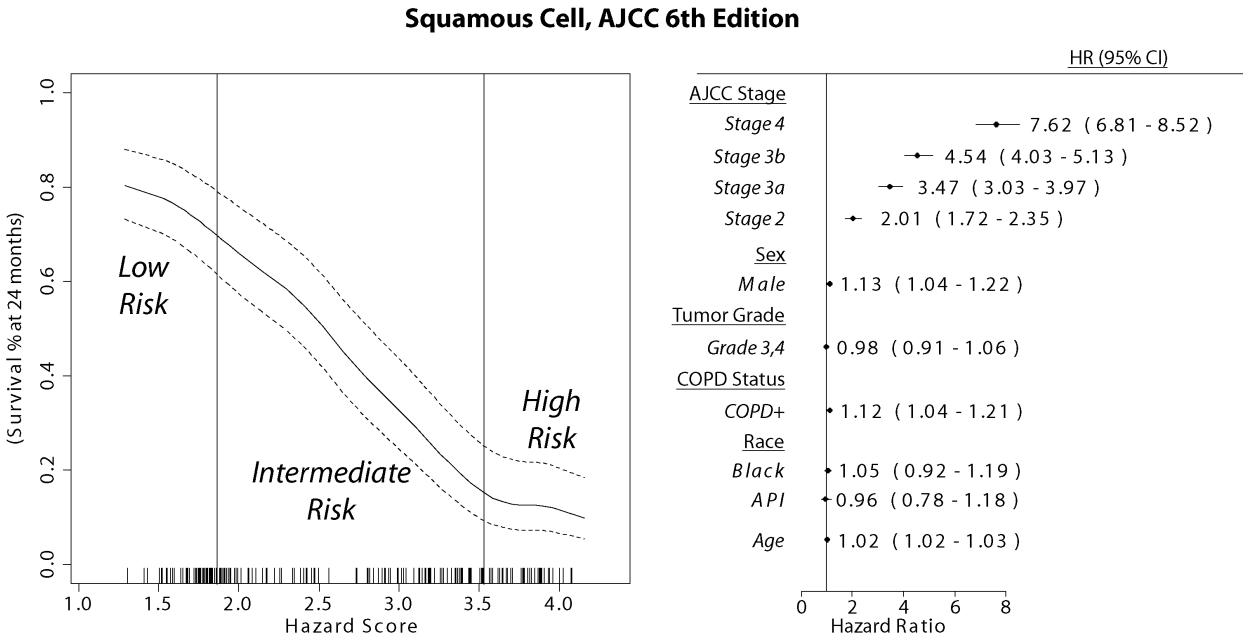

Figure 3.13. Final model for Squamous Cell AJCC 6th Edition. Model coefficients used to determine the Hazard Score for each patient are shown on the forest plot (right). Patient survival at 24 months for the total population sample is shown for the range of Hazard Scores (left), with the risk-groups delimited by vertical bars.

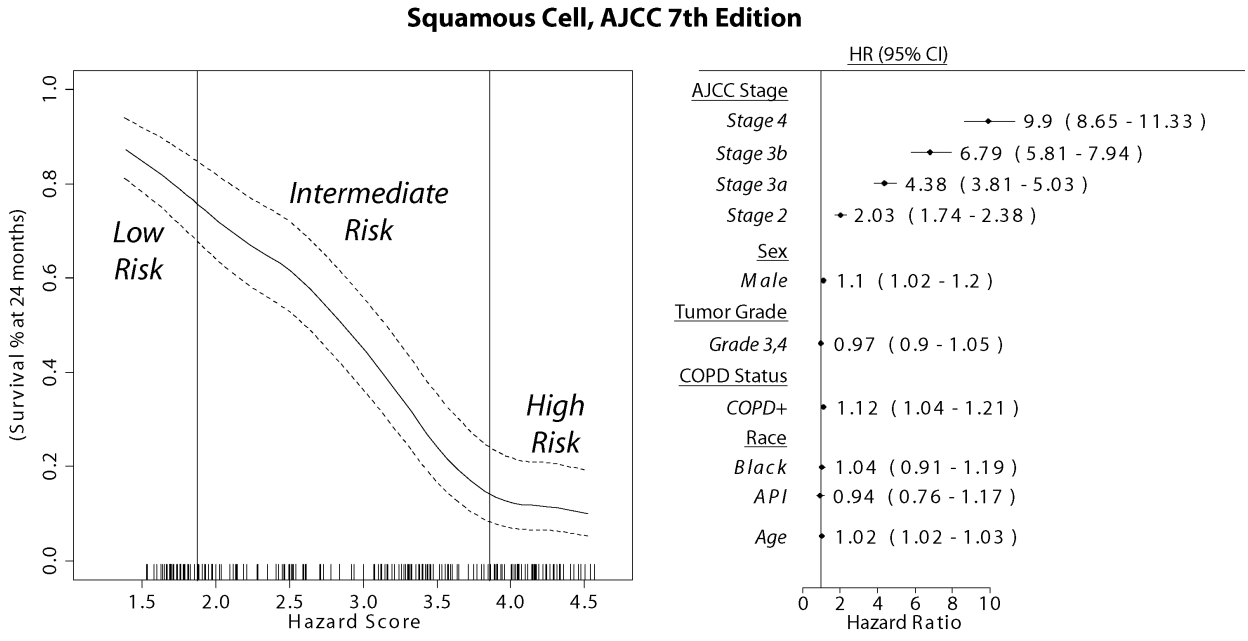

Figure 3.14. Final model for Squamous Cell AJCC 7th Edition. Model coefficients used to determine the Hazard Score for each patient are shown on the forest plot (right). Patient survival at 24 months for the total population sample is shown for the range of Hazard Scores (left), with the risk-groups delimited by vertical bars.

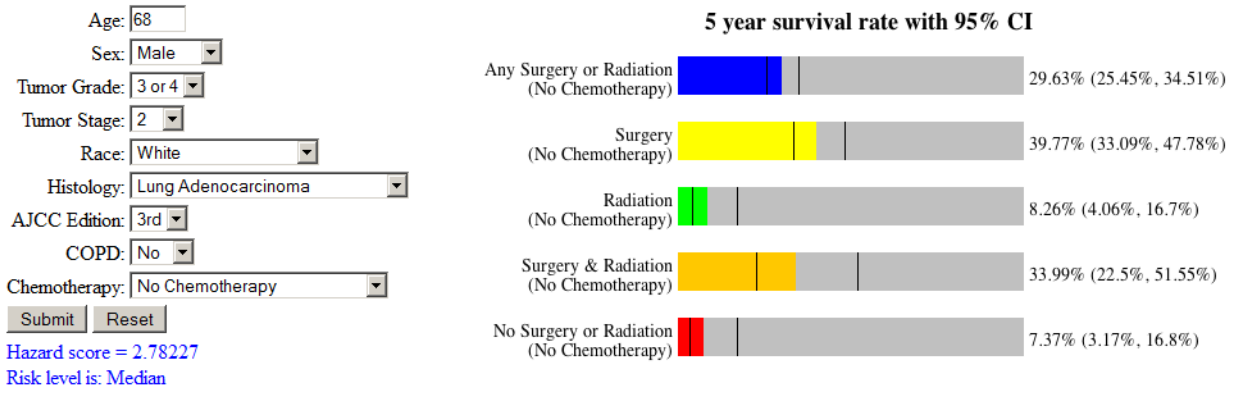

Figure 3.15. Sample output from the web-based version of the final model. Given the patient values submitted by the user (left), the web-based model will estimate survival for each treatment category using the survival observed for patients of a particular treatment modality and similar Hazard Score (right).

Discussion

It was shown that co-morbidities derived from administrative data contribute to a significant improvement in prognostication. Specifically, the inclusion of a COPD variable was able to significantly improve prognostication in several of the treatment groups considered as measured by metrics of model fit and accuracy in controlled analyses and as an independent predictor. The presence of COPD appears to have a strong effect on survival across the length of follow-up, with COPD patients having significantly worse survival curves in four of the six groups considered. This was reflected in the proportional test of COPD prevalence, where the group of patients with relatively shorter survival had a significantly higher prevalence of COPD. Because a COPD designation was based on treatment of the disease, the increased prevalence in patients with relatively shorter survival may reflect a higher degree of COPD severity as a result of more severe cases having a higher likelihood of clinical diagnosis and treatment. However, a quantitative determination of that effect is outside the scope of this analysis.

Additionally, splitting the patients further by the type of chemotherapy administered did not reduce the prognostic ability of the covariate model previously described when compared to a model using tumor stage alone(Putila et al., 2011), indicating that variable administration of chemotherapy does not confound the benefit seen from the inclusion of other pathologic and demographic variables. The relatively stronger effect of COPD in the squamous cell carcinoma patients in models controlling for other variables may be indicative of the role of smoking on both survival and the presence of COPD itself, given differences in the proportion of smokers between histologies (Papi et al., 2004; Kenfield et

al., 2006; Bryant & Cerfolio, 2007), with squamous cell carcinomas being more closely associated with smoking than adenocarcinomas.

Assessment of the effect of COPD is also important due to potential shared origins in the development of the disease (Engels, 2008; Yao & Rahman, 2009; Houghton et al., 2008) and its role as a potential confounder (Young et al., 2008) in prognostic models. It is hypothesized that inflammation may initiate or promote tumorigenesis in the lung. This action is thought to center around the induction of immune dysfunction and the destruction of the extra-cellular matrix. While it is difficult to demonstrate quantitatively given the data available in this set, there was no evidence found which would suggest that the trajectory or treatment of COPD after diagnosis will affect survival estimation in lung cancer patients when using disease-specific survival. It is however necessary to examine the role of COPD within treatment and stage sub-set, in order to account for known confounders such as treatment candidacy and complication rate (Janssen-Heijnen et al., 1998; Rancati et al., 2003). After considering the effects of COPD in influencing treatment selection, there is no compelling statistical reason to avoid the use of a proportional hazards model or COPD status despite possible shared origins between COPD and lung cancer *in vivo*. Conversely, other co-morbid conditions such as heart or cerebro-vascular disease are more likely to derive their prognostic value from confounding or artifactual influences such as socio-economic status or lifestyle differences, as there are only tenuous causal pathways for their effects on disease-specific survival. Thus, their inclusion, while potentially beneficial from a statistical standpoint, is imprudent when attempting to model factors directly influencing patient survival. For these reasons, and as a matter of simplicity and generalizability, COPD alone was added to the previous model in lieu of multiple conditions.

The SEER portion of the data is composed of a geographically and demographically diverse group of patients by design and thus allows for a more accurate analysis of the effect of multiple variables on survival. The large size and scope of the data also allow for assessing the accuracy of the model across a wide range of treatments and patient profiles. This is especially relevant when studying the effects of race or comorbidity, factors which have previously been shown to influence survival due to differences in treatments administered (Bach et al., 1999). Given the assessment of disease-specific survival stratified by treatment modality, it is highly likely that the effect of COPD represents an additional source risk to the patient beyond that attributable to reduced candidacy, complication rate, or reduced lung function after treatment. The model also accounts for recoding of cancer cases to the newest version of the AJCC staging schema, where possible, and includes both major non-small cell

histologies although the utility of COPD is less pronounced in the 6th Edition and recoded 7th Edition samples. A possible explanation for this is that the newer patient groups have both a smaller sample-size and a shorter period of follow-up. It is possible that important manifestations of differences in survival due to COPD may occur beyond the three-year mark or that more recent improvements in the clinical management of COPD cases may alleviate some of the disparity in survival, but this was not assessed.

A limitation of the data is that Medicare coverage is limited to those over the age of 65, with some exceptions, and thus the age distribution of patients included is likely biased toward the upper end of what would be seen in clinical practice. Additionally, the administration of chemotherapeutics or other treatments not covered by Medicare is not recorded in the data, although Medicare covers a significant portion of cancer care in patients who are eligible (Thorpe & Howard, 2003). As the treatments administered to each patient were non-random, it is also difficult to accurately assess the relative benefit of each type of treatment with a high degree of certainty and thus for this and other reasons the results should not be interpreted as a measure of comparative treatment effectiveness. It should also be noted that although administrative data is adequate for assessing co-morbidities and the use of chemotherapeutics (Deyo et al., 1992; DHoore et al., 1996), these ascertainments may have limited generalizability to more refined clinical or functional measures capturing the severity or duration of disease. The absence of information on smoking status may also limit the model, as smoking has been seen to influence survival independent of co-morbidities (Tammemagi et al., 2004). While the metrics used to assess model performance indicate that the proposed model has a high degree of internal validity, an external validation set was not used and thus caution must be exercised when generalizing to other populations. Conversely, this type of prognostic model shows the greatest individual-level improvement in survival when identifying high-risk patients who would otherwise have been misclassified and treated as a lower-risk patient when using stage alone. In that regard, the assessments used bias the results toward the null, as potential large improvements in a single patient are averaged over the group.

Our results suggest that the addition of co-morbid conditions, specifically COPD, to a comprehensive prognostic model improves prognostication over similar models without that information. This represents an additional improvement over the use of stage alone, even in a statistically rigorous manner, which was shown previously. Additionally, stratifying patients by the administration of chemotherapeutic agents does not affect the performance of the model. The

presence of COPD in patients may have an effect on survival independent of the effect attributable to disparate treatment candidacy or quality of life. Thus, this approach and the resulting model have value in identifying high-risk patients who may benefit from more aggressive therapies or enhanced follow-up.

Chapter4:
Projection of prognostic and chemo-predictive microRNA markers from a squamous cell lung cancer cohort to linked SEER-Medicare data

Joseph Putila, Nancy Lan Guo

Abstract

Introduction: MiRNA are potentially important moderators of chemoresponse in human cancer due to their widespread regulatory function. Identification of a set of clinically validated and biologically relevant miRNA could serve as potential markers for prognosis, chemoresponse prediction, and novel therapeutic targets. The objective of this study was to utilize population, clinical, and cell-line data as an integrative approach to identifying miRNA showing a consistent association with chemoresponse.

Materials and Methods: Patients diagnosed with squamous cell carcinomas of the lung between 1991 and 2005 were retrieved from the linked SEER-Medicare database with data on tumor TNM markers, grade, survival, and treatment information (N=33,897). Medicare HCPCS and Revenue codes were used to identify patients receiving cisplatin (N=873), carboplatin (N=4,267), etoposide, (N=318), and paclitaxel (N=3,098). The AJCC TNM markers and grade were used to partition patients into disjoint groups indicative of tumor progression, with the average survival being calculated for each group. A second set of clinical data containing miRNA expression (N=57, GSE 16025) was partitioned in the same manner, with significant miRNA being selected by Cox and Kaplan-Meier modeling. Linear regression was used to estimate the association between average miRNA expression in the clinical cohort and survival in the population sample treated without chemotherapy across groups with similar tumor characteristics. The same analysis was performed on the population set of patients receiving one of the four chemotherapies, and compared to the results using the clinical cohort. A third analysis utilizing the NCI-60 anti-cancer screen estimated the association between miRNA expression and drug activity in an experimental setting. MiRNA which showed a significant association in the population set and concordance in all others were considered for further analysis using IPA software.

Results: Multiple miRNA, including miR-433 and miR-520d* showed consistent associations with survival across analytic approaches and treatments in the non-chemotherapy group. In patients treated with chemotherapy, miR-199b and miR-142-3' showed associations with chemoresponse in the clinical and

cellular analyses. These and other significant miRNA formed networks which included experimentally validated links with mediators of tumor growth, apoptosis, and cell-cycle control.

Introduction

Lung cancer is currently the leading cause of cancer-related death in the U.S., due in part to the minimal response to chemotherapy and potential for recurrence (Hoffman et al., 2000; Naruke et al., 1989; Vogt et al., 1999). Patients with similar clinical and pathological profiles may have markedly different responses to a given treatment, indicating that classical markers of progression or tumor type alone are not sufficient for the selection of an appropriate course of therapy. The use of biomarkers to enhance treatment selection is a promising avenue for improving patient outcomes.

Numerous studies have shown that mRNA biomarkers can be used for predicting survival in lung cancer patients (Wan et al., 2012; Raponi et al., 2006; Chen et al., 2007; Beer et al., 2002). Similar studies have shown that miRNA can also be used in the diagnosis and prognosis of lung cancer (Wang et al., 2009; Rabinowits et al., 2009; Yu et al., 2008; Markou et al., 2008; Fabbri et al., 2007; Raponi et al., 2009). The development of a biomarker model of chemoresponse is particularly valuable due to the high degree of chemoresistance in lung cancer (Vogt et al., 1999). Such a model has the potential to both enhance response rates seen in patients treated with chemotherapy and to highlight the potential regulatory networks which influence patient survival and mediate chemoresponse.

The variation in response to chemotherapy between patients is likely mediated, at least in part, by micro-RNA (miRNA). These miRNA are short non-coding RNA sequences which have been shown to be promising markers for both prognosis and response to therapeutics in multiple cancer types (Raponi et al., 2009; Yan et al., 2011; Yu et al., 2008). There are a number of advantages to using miRNA such as their presence in circulating plasma (Chen et al., 2012; Mitchell et al., 2008), and greater stability in prepared tissue samples relative to mRNA (Jung et al., 2010; Mraz et al., 2009) including formalin fixation (Xi et al., 2007). The use of miRNA markers in the selection of a biologically appropriate treatment has the possibility of improving patient outcomes by determining the best application of existing drugs, and the elucidation of pathways which may aid in the development of novel treatments (Iorio & Croce, 2012; Avraham & Yarden, 2012).

In selecting miRNA targets, many approaches have limited power to consider the effects of surgical or radiological treatment due to sample size issues prompted by cost and recruitment barriers. This problem is compounded by assignment to differing treatments based on traditional clinical markers, confounding the use of survival time as an outcome. One potential solution is to utilize data from larger clinical populations found in cancer registries to model the effects of miRNA expression on patient outcomes in groups with similar surgical and radiological treatments. This approach allows for the selection of miRNA markers which have promise for prognostication and prediction in a large and diverse patient population. Similar approaches have had success identifying genes associated with patient survival based on aggregate measures taken from population data (Stein et al., 2004). Additionally, the approach more appropriately takes into account the effects of surgical and radiological treatments in combination when assessing the role of miRNA in mediating chemoresponse as represented by patient survival. Careful analysis of the known, experimentally validated biological networks provides insight into the cellular processes which may be mediating this response, and several interaction networks are presented as potential mediators of chemoresponse.

Materials and Methods

Patient cohorts

Two separate groups of patients were used in the analysis. The first is a cohort of 57 squamous cell carcinoma patients originally published by Raponi et al. (Raponi et al., 2009). Included in this cohort is expression data on 328 human miRNA determined by MirVana miRNA Bioarrays, with corresponding follow-up information on survival time and status, tumor grade, and AJCC tumor T, N, and M markers. Tumor grade for this cohort was converted from a descriptive measure of tumor differentiation to numerical grade to match descriptors used in the second set.

The second patient group was derived from the linked SEER-Medicare (SM) database (Warren et al., 2002b). This database combines the clinical, demographic, pathological, and survival information from the NCI SEER registry system with claims data for individual patients found in the Medicare claims database. Briefly, patients in the SEER portion were linked according to identifying information such as social security number, census tract, age, and other identifying criteria to records found in the Medicare claims database (Potosky et al., 1993). Criteria for inclusion in this study were a diagnosis of squamous cell carcinoma of the lung or bronchus between 1991 and 2005, as well as valid information on survival

time and status, tumor grade, and AJCC T, N, and M markers. A total of 33,897 patients fit these criteria, with this sample being stratified according to treatment modality in subsequent analyses. Patient demographics for this group are detailed in Table 4.1. The T, N, and M markers for patients diagnosed prior to 2004 were derived from EOD10 coding, where possible, and according to established conversion algorithms (Seiffert, 1993). Four specific chemotherapeutic agents, cisplatin, carboplatin, paclitaxel, and etoposide were considered. The administration of chemotherapy was determined through the use of Healthcare Common Procedure Coding System (HCPCS) codes (Warren et al., 2002a). First, the use of chemotherapy was determined by searching individual patient claims histories for entries with an HCPCS code corresponding to the agent in question. The ICD-9 diagnosis codes for these records were then checked to ensure that the agent was administered for the treatment of lung cancer. Curative surgery and radiation therapy were determined using variables in the SEER portion of the data. These variables were used to stratify patients into a group receiving any surgical procedure but not pre-operative radiation, a group with any type of radiation but not surgery, a group with both surgery and radiation, and lastly a group with any combination of treatments. Survival estimates are represented as disease-specific survival for the SM cohort and overall for the clinical cohort.

Table 4.1. Summary of demographic and clinical variables for patients in the linked SEER-Medicare and Clinical cohorts.

Variable	SEER-Medicare Patients	Clinical Cohort (Raponi et al)
Cancer Stage		
Stage I	12,651 (37.3%)	34 (59.6%)
Stage II	2,662 (7.8%)	11 (19.3%)
Stage III	11,514 (34%)	12 (21.1%)
Stage IV	5,813 (17.1%)	0 (0.0%)
Unstaged/Other	1,257 (3.7%)	0 (0.0%)
Tumor Grade		
Grade 1	1,462 (4.3%)	1 (1.8%)
Grade 2	13,573 (40.0%)	28 (49.1%)
Grade 3	18,202 (53.7%)	28 (49.1%)
Grade 4	660 (1.9%)	0 (0.0%)
Mean Age (σ)	71.1 (7.9)	66.8 (10.7)
Sex		
Male	22,218 (65.5%)	39 (68.4%)
Female	11,679 (34.5%)	18 (31.6%)

Cellular Data

The cellular data against which the predictive markers are checked was derived from the NCI-60 cell panel (Shoemaker, 2006; Gaur et al., 2007) and contains miRNA expression levels of 209 miRNA markers determined by quantitative-PCR across 59 cancer-derived cell lines of diverse tissue origin. The data also contain the drug activity at three clinically relevant end-points: 50% Growth Inhibition (GI50), Total Growth Inhibition (TGI), and 50% Lethal Concentration (LC50), used here to refine the set of significant miRNA and assess meaningful biological context.

Identification of prognostic miRNA using a clinical cohort

Patients in the Raponi and SEER-Medicare sets were assigned to disjoint tumor progression groups according to unique combinations of tumor grade and T, N, and M markers. This tumor progression group membership served as a link between the two patient sets in subsequent analyses. In total, there were 16 groups common between the clinical and population data sets, including tumor grades 1 through 3, tumor T 1 through 4, and tumor N 0 through 2. There were no metastatic or grade 4 groups common between the two sets. A Leverage analysis was performed to test for undue influence from any one group. None of the groups were seen to have a disproportionate effect on model coefficients or significance; therefore all 16 groups were included.

The set of prognostic miRNA was determined by selecting miRNA which showed a significant association with survival in the original clinical cohort, then validating these miRNA in the population cohort. Cox modeling and Kaplan-Meier estimation were used to assess the association between expression and survival in the original clinical cohort. Cox model coefficients and p-values were estimated for each miRNA in independent models. In the Kaplan-Meier analysis, cutoff values ranging from the 5% to 95% quantile were used to split the patients into high and low survival groups. The degree of separation between the resulting survival curves was estimated as a log-rank p-value, with a value less than 0.05 being deemed as significant separation between the groups.

Validation of prognostic miRNA markers using linked SEER-Medicare data

Validation of the results of the clinical analysis on the population data was done using linear regression, Cox modeling, and Kaplan-Meier estimation. Multiple methods of assessment were chosen due to limitations associated with using a single measure (Pepe et al., 2004). The linear regressions used the average miRNA expression from the clinical set and average disease-specific survival from the

population set for each tumor progression group. Average survival in the population cohort was calculated as a function of the area under the curve produced by Kaplan-Meier estimation. Each miRNA was evaluated in each of the four surgical and radiological treatment modalities. In order to enter the final set of prognostic miRNA, each miRNA found to be significant in the original cohort had to show a concordant and significant association with average survival in the population cohort. Individual patients in the population cohort were then assigned the average miRNA expression for that progression group from the clinical cohort and the Cox and Kaplan-Meier models were re-evaluated. Any miRNA which failed to achieve a significant and concordant association with survival in one of either the Cox or Kaplan-Meier models were removed.

Identification of predictive miRNA using linked SEER-Medicare data

In order to select for miRNA which were predictive of chemoresponse, as represented by improved or diminished disease-specific survival, linear regression was again used to estimate the association between average miRNA expression in the clinical cohort and disease-specific survival in the population cohort by tumor progression group. Next, each patient in the population set was assigned the average expression for each miRNA corresponding to the same tumor progression group in the clinical cohort. This combined set of patient expression and survival data was then used to estimate a Cox proportional hazards model. A Kaplan-Meier model was also estimated and assessed on the log-rank P-values. Significance and concordance on the linear model and one of either the Cox or Kaplan-Meier models was sufficient evidence in the population analysis. Selection of a particular miRNA as a prognostic marker was not a requirement for selection as a predictive marker.

Validation of predictive miRNA using the NCI-60 anti-cancer screen

The miRNA which were significant in the predictive analyses were then compared to data from the NCI-60 anti-cancer screen (Boyd & Pauli, 1995; Gaur et al., 2007; Weinstein, 2006). Linear regression was used to estimate the association between expression on a specific miRNA marker in each cell line and drug activity in the same cell line for each of the previously described end-points. Cell-lines which did not have informative values, specifically those which did not achieve the end-point in question at any dosage, were removed from the analysis. Significance and concordance on any one of the three drug activity measures was considered support of meaningful biological context, and these miRNA formed the final of predictive set.

Evaluating Predictive miRNA using Pathways Analysis

For each agent, the final set of significant miRNA were also examined for interactions with molecular species known to play a role in lung cancer or relevant cellular processes such as apoptosis, proliferation, cell-cycle regulation, or metastasis through the use of Ingenuity Pathways Analysis (IPA) (Ingenuity® Systems, www.ingenuity.com). In short, IPA is a functional pathway analysis tool incorporating genes, cellular species such as proteins, and chemical compounds with data on their interactions and involvement in diseases derived from scholarly publications. Using this data, it is possible to map interactions between biomarkers on any given criterion. The list of significant miRNA from each treatment was used to create networks based on experimentally validated interactions between miRNA and molecular components with known biological function. These networks were created using the Core Analysis feature of IPA, and can be seen in Figures 4.5 through 4.8. The Core Analysis compares the set of miRNA markers with molecules with known roles in human disease in order to select a set of networks in which the interactions between the miRNA set and IPA-defined functional set are statistically over-represented.

Results

Prognostic miRNA

Prognostic analyses were performed in the group of patients without indication of having received chemotherapy. In patients treated with surgery and without chemotherapy miR-520d*, miR-433, miR-134, and miR-382 were positively and three others were negatively associated with survival. In the sub-set of patients receiving radiation without surgery or chemotherapy, a single miRNA, miR-433 was positively associated with survival. In patients receiving both surgery and radiation without chemotherapy, there were 3 positive and 1 negative associations, with miR-520d* again being positively associated with survival. For patients receiving any surgical or radiation treatment without chemotherapy, 9 miRNA including miR-520d* and miR-433 were positively associated, with miR-384 and two others negatively associated with survival. The specific miRNA and their relationship with disease-specific survival are outlined in Table 4.2. Multiple miRNA were significant across surgical and radiation modalities. Two miRNA, miR-433 and miR-520d*, were positively associated in multiple treatment combinations considered, indicating a strong but non-specific effect on survival. The results of the analyses on miR-433 and miR-520d* are detailed in Figures 4.1 and 4.2, respectively.

Table 4.2. Total number of significant miRNA when considering variable administration of surgery, chemotherapy, and radiation in the SEER-Medicare and clinical cohort.

Treatment	No Chemotherapy	
	Positive Correlation	Negative Correlation
Surgery	miR-520d*, miR-433, miR-134, miR-382	miR-328, miR-384, miR-525*
Radiation	miR-433	-
Surgery and Radiation	miR-453, miR-520d*, miR-134	miR-197
Any Treatment	miR-453, miR-372, miR-142-3', miR-329, miR-520d*, miR-433, miR-134, miR-382, miR-493-3'	miR-328, miR-197, miR-384

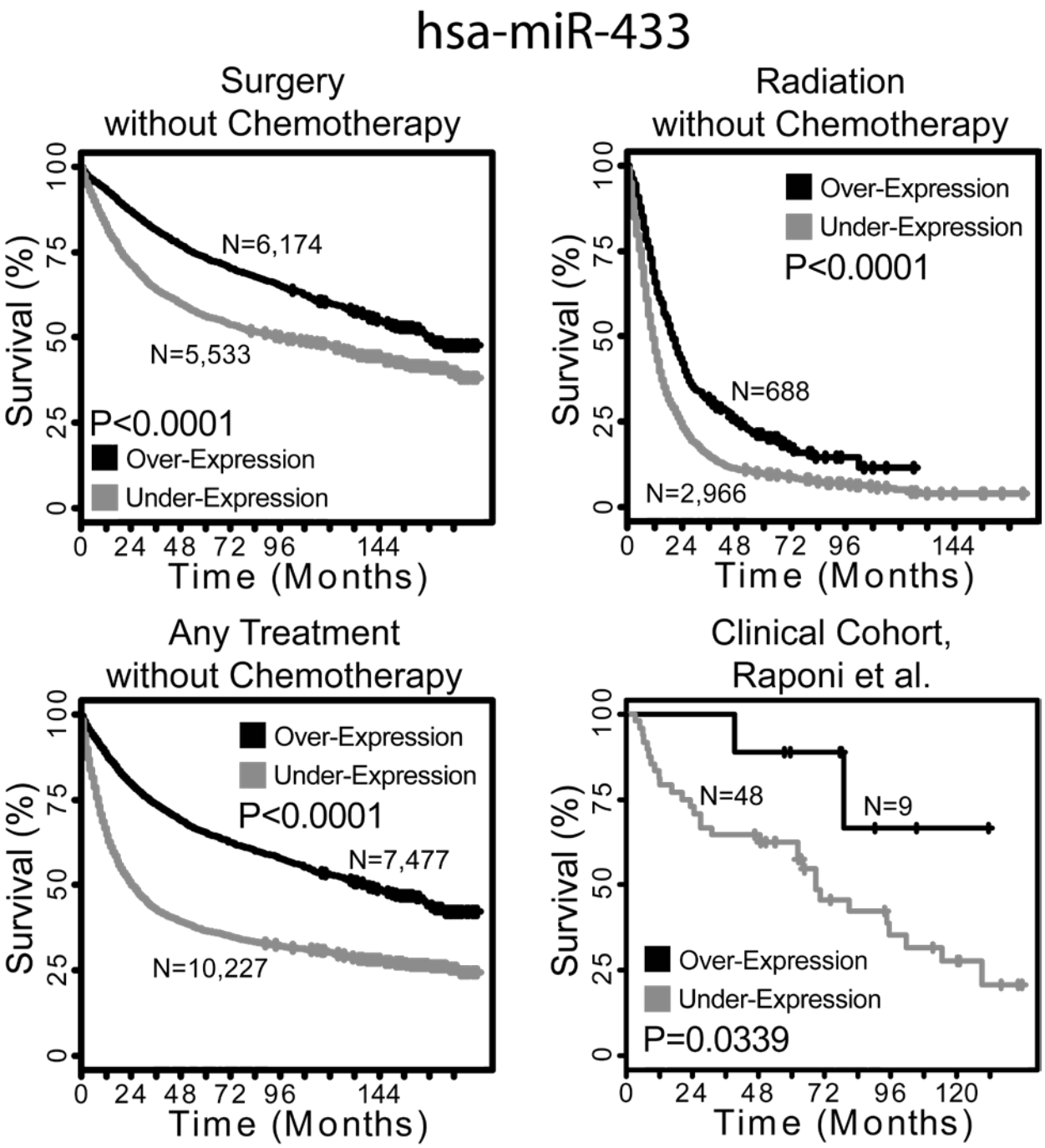

Figure 4.1. Relationship between expression of miR-433 and survival in patients from the population set treated without chemotherapy, and the original clinical cohort. Treatment groups shown are surgery without chemotherapy (top left), radiation without chemotherapy (top right), any treatment without chemotherapy (bottom left), and the clinical cohort regardless of treatment (bottom right).

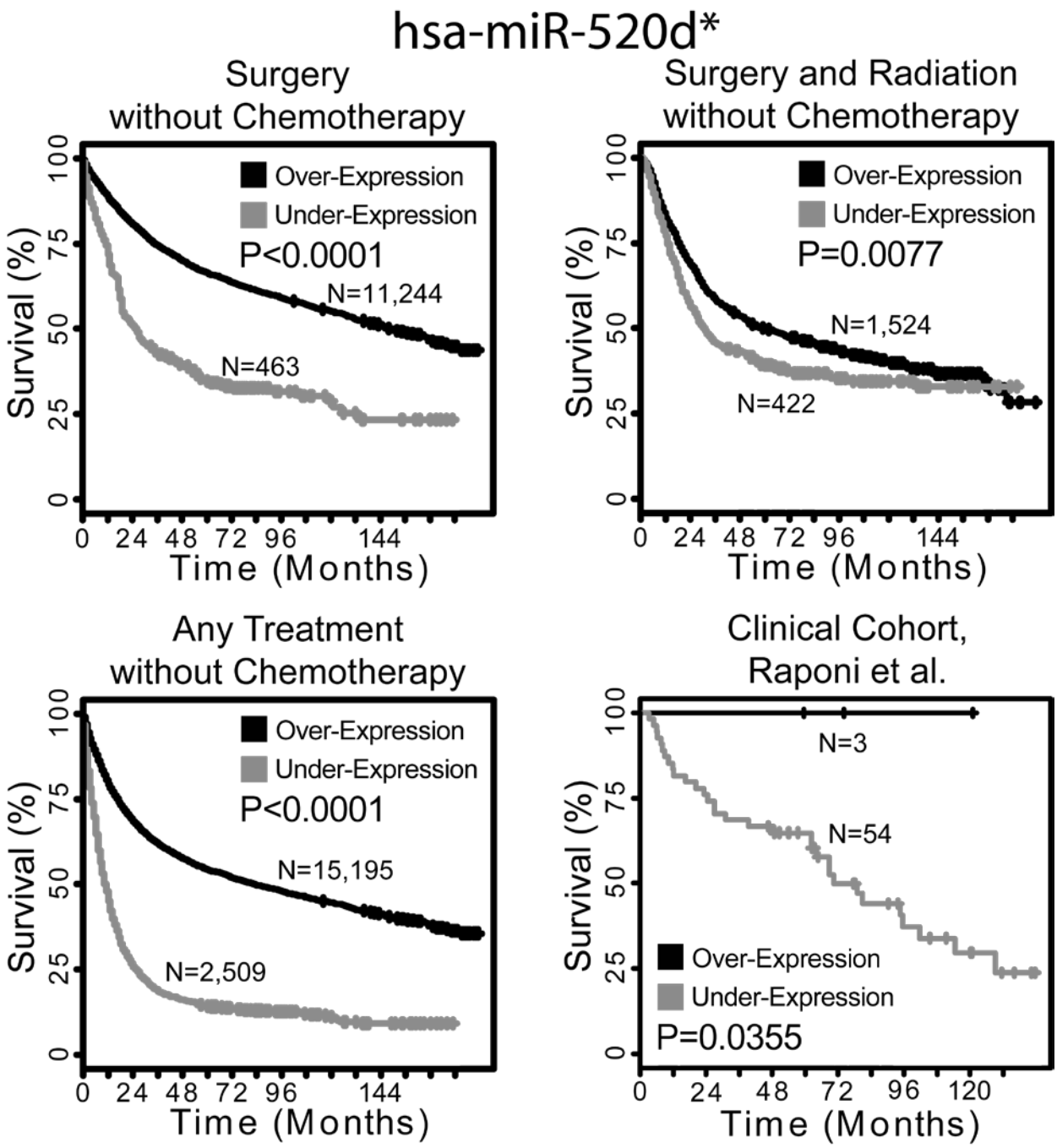

Figure 4.2. Relationship between expression of miR-520d* and survival in patients from the population set treated without chemotherapy, and the original clinical cohort. Treatment groups shown are surgery without chemotherapy (top left), both surgery and radiation without chemotherapy (top right), any treatment without chemotherapy (bottom left), and the clinical cohort regardless of treatment (bottom right).

Predictive miRNA

In addition to those found to be prognostic, numerous miRNA were found to predictive of survival when patients treated with any chemotherapy or a specific agent were considered independent of those patients not receiving chemotherapy. The results of these analyses are summarized in Tables 4.3 and 4.4, with the significant sets being split into chemo-sensitive and chemo-resistant for presentation.

Cisplatin

In patients who received surgery with cisplatin, 9 miRNA were negatively associated, with one miRNA positively associated. Two of the negatively associated miRNA in that group, miR-181b and miR-181c have highly similar mature sequences. For patients receiving radiation with cisplatin a single miRNA, miR-384, was negatively associated. In patients receiving both surgery and radiation there were no significant associations in the cisplatin group. For those receiving cisplatin in combination with any surgical or radiation treatment there were 8 positive and 8 negative associations.

Carboplatin

When patients receiving surgery with carboplatin were considered miR-134, miR-142-3', and two others were positively associated. An additional 4 were negatively associated. In patients receiving radiation with carboplatin there were not any significant associations. In patients receiving both surgery and radiation miR-134 and miR-142-3'were positively associated and 2 others negatively associated. When patients receiving carboplatin in combination with any treatment were considered, a total of 6 miRNA were positively associated with survival, with a further 4 miRNA being negatively associated. Tables 4.3 and 4.4 detail the identities of the significant miRNA by treatment modality.

Paclitaxel

In patients receiving paclitaxel with surgery there were 5 miRNA positively and 7 negatively associated with survival. Two miRNA, miR-154 and miR-302a*, were positively associated with survival in patients receiving radiation with paclitaxel with one negative association. In patients who received both surgery and radiation miR-142-3' and miR-220 were positively associated with 5 others negatively so. In patients receiving paclitaxel in combination with any other treatment there were 5 miRNA positively associated and 6 miRNA negatively associated with survival. Both miR-142-3' and miR-199b

were also significant in the clinical and cellular analyses, the results of which are outlined in Figures 4.3 and 4.4, respectively.

Etoposide

For patients receiving etoposide in combination with any treatment, 9 miRNA were positively and 9 negatively associated with survival. There were no miRNA significantly associated with survival in patients receiving only surgery. A single miRNA, miR-223 was negatively associated with survival in patients receiving both surgery and radiation. For patients receiving radiation with etoposide, there were 9 positive and 5 negative associations.

Non-specific Chemotherapy

The last group considered consisted of patients with an indication of having received any chemotherapy, including any of the four previously listed. Patients in this group having received surgery with chemotherapy had 2 positive and 2 negative associations, including miR-142-3' being positively associated as was the case for the specific treatment groups. In the radiation group there was 1 positive and 1 negative association. In the group receiving both surgery and radiation there was 1 positive association, miR-142-3', along with 2 negative associations. When any treatment with chemotherapy was considered, there were 14 positive associations, including both miR-520d* and miR-433, and 7 negative associations.

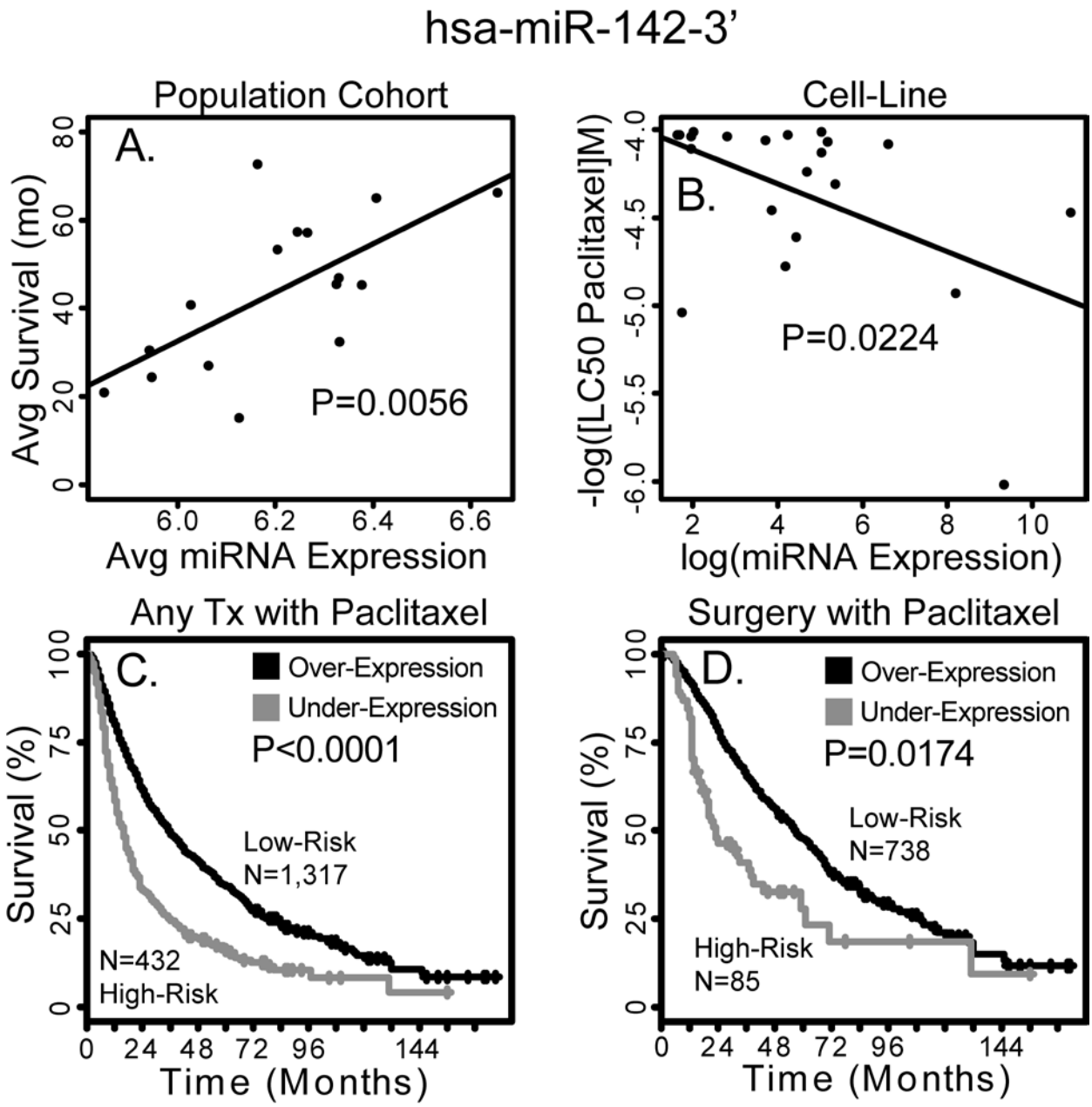

Figure 4.3. Relationship between expression of miR-142-3' and survival in the population analysis, and expression and LC50 in the cellular panel when treated with paclitaxel. Average expression of miR-142-3' showed a positive correlation with survival in the population analysis (A), which was corroborated in the cellular panel by a decrease in the dosage required for LC50 (B). This effect was also apparent when stratifying the groups based on expression level in all patients treated with paclitaxel (C), and persisted when only the surgical group was considered (D).

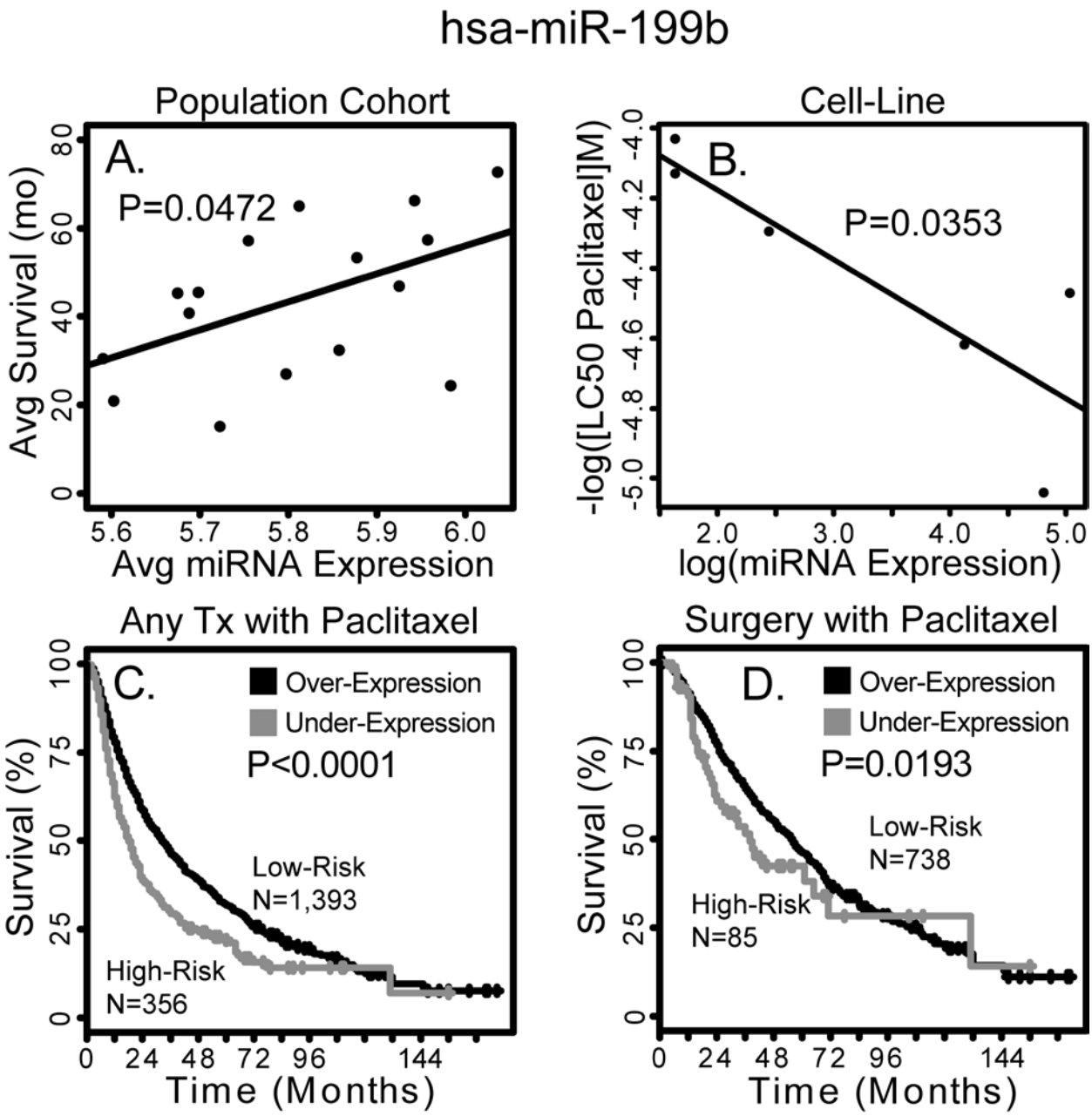

Figure 4.4. Relationship between expression of miR-199b and survival in the population analysis, and expression and LC50 in the cellular panel when treated with paclitaxel. Average expression of miR-199b showed a positive correlation with survival in the population analysis (A), which was corroborated in the cellular panel by a decrease in the dosage required for LC50 (B). This effect was also apparent when stratifying the groups based on expression level in all patients treated with paclitaxel (C), and persisted when only the surgical group was considered (D).

Table 4.3. Total number chemosensitive miRNA when considering administration of specific chemotherapeutic agents in combination with variable administration of surgery and radiation, as evidenced by prolonged disease-specific survival. The miRNA which produced a significant stratification in the SEER population samples in addition to significance on one or more measures of drug activity in the NCI-60 panel are bolded.

Treatment	Cisplatin	Carboplatin	Paclitaxel	Etoposide	Any Chemo
Surgery	miR-29b	miR-134, miR-142-3', miR-188, miR-380-5'	miR-134, miR-138, **miR-142-3'**, miR-188, miR-380-5'	-	miR-142-3', miR-433
Radiation	-	-	miR-154, miR-302a*	miR-141, miR-17-3', miR-17-5', miR-182, miR-183, miR-19b, miR-200c, miR-222, miR-23b	miR-134
Surgery and Radiation	-	miR-134, miR-142-3'	**miR-142-3**', miR-220	-	miR-142-3'
Any Treatment	miR-129, miR-134, miR-142-3', miR-184, miR-198, miR-370, miR-373, miR-379	miR-134, miR-142-3', miR-206, miR-33, miR-370, miR-372	miR-134, **miR-142-3'**, **miR-199b**, miR-370, **miR-382**	miR-129, miR-141, miR-142-3', miR-184, miR-218, miR-220, miR-335, miR-373, miR-96	miR-33, miR-453, miR-372, miR-142-3', mir-299-3', miR-329, miR-520d*, miR-519a, miR-494, miR-433, miR-134, miR-485-5', miR-518c*, miR-493-3'

Table 4.4. Total number chemoresistant miRNA when considering administration of specific chemotherapeutic agents in combination with variable administration of surgery and radiation, as evidenced by shortened disease-specific survival. The miRNA which produced a significant stratification in the SEER population samples in addition to significance on one or more measures of drug activity in the NCI-60 panel are bolded.

Treatment	Cisplatin	Carboplatin	Paclitaxel	Etoposide	Any Chemo
Surgery	miR-126*, miR-136, miR-181b, **miR-181c**, miR-196a, **miR-331**, **miR-375**, miR-424, **miR-92**	miR-126, **miR-192**, miR-195, miR-384	miR-189, miR-192, miR-197, miR-301, miR-328, miR-331, miR-384	-	miR-197, miR-384
Radiation	miR-384	-	miR-189	miR-150, miR-155, miR-192, miR-337, miR-98	miR-328
Surgery and Radiation	-	**miR-192**, miR-384	miR-192, miR-197, miR-328, miR-331, miR-384	miR-223	miR-328, miR-197, miR-384
Any Treatment	miR-132, miR-181b, miR-30c, miR-30e-3', miR-324-3', **miR-331**, miR-339, miR-384	miR-126, miR-195, miR-224, miR-384	miR-126, miR-224, miR-328, miR-331, miR-361, miR-384	miR-132, miR-133b, miR-155, miR-197, miR-208, miR-214, miR-324-3', miR-374, miR-423	miR-328, miR-361, miR-511, miR-197, miR-125a, miR-384, miR-126

Discussion

These results demonstrate that, based on similarities in tumor progression, extrapolation of miRNA expression from smaller cohorts to larger population-based data can serve both as an additional confirmatory tool where novel cohorts are unavailable and as an independent source of prognostic and predictive markers. Specifically, it was shown that multiple miRNA were selected as strong predictors of chemoresponse though analysis of disease specific survival in cohorts with similar treatment strategies. The miRNA selected for also have a strong biological context relative to cellular processes important to cancer progression and patient survival such as apoptosis, suggesting that they play a role in mediating the effectiveness of chemotherapeutic agents through their regulatory function.

In patients treated with paclitaxel, miR-199b was positively associated with survival, an effect which may be indicative of a number of potential regulatory actions. In particular, miRNA-199b may decrease translation of LAMC2, a key growth factor for non-small cell lung cancer, and may confer a protective effect due to a reduction in metastatic potential (Manda et al., 2000) and loss of protection against anoikis (Kodama et al., 2005). The same miRNA also increases inhibition of PAK4, a protein kinase suspected of promoting tumor survival, proliferation, and metastasis in a variety of cancers (Mak et al., 2011; Liu et al., 2008). Similarly, miR-29b was seen to be protective and may also reduce metastatic potential by acting as a DNMT inhibitor and reverting aberrant methylation patterns (Chang et al., 2006; Fabbri et al., 2007). The miRNA in the significant sets also had known associations with one or more functional networks pertinent to chemoresponse. Additionally, cancer was the top disease for three of four specific chemotherapy groups considered based on the degree of overlap between the significant miRNA and molecules identified in IPA as being related to cancer. A summary of these networks is provided in Tables 4.5 through 4.7. The networks resulting from the Core Analysis performed in IPA are shown in Figures 4.5 through 4.8.

Table 4.5. Networks derived from a *Core Analysis* performed in Ingenuity Pathway Analysis. The corresponding networks are detailed in Figures 4.5 through 4.8. Focus molecules represent miRNA from the original set which demonstrated a high degree of interconnectivity, and were subsequently used to establish the network score. A significant p-value indicates a low probability of the miRNA set being selected at random.

Treatment Group	Network ID	Focus Molecules	p-value
Cisplatin	Connective Tissue Disorders, Inflammatory Disease, Inflammatory Response	14	<0.0001
Carboplatin	Connective Tissue Disorders, Inflammatory Disease, Inflammatory Response	11	<0.0001
Paclitaxel	Genetic Disorder, Skeletal and Muscular Disorders, Developmental Disorder	17	<0.0001
Etoposide	Respiratory Disease, Inflammatory Response, Connective Tissue Disorders	14	<0.0001

Table 4.6. Top Diseases and Disorders derived from a *Core Analysis* performed in Ingenuity Pathway Analysis. The corresponding diseases or disorders were selected based on the degree of overlap between the miRNA selected in that treatment group and those defined in the IPA database as being associated with that disease beyond random chance ($p<0.05$). These results, unlike those represented in Table 4.5, do not factor the degree of connectivity.

Treatment Group	Disease or Disorder	# Molecules
Cisplatin	Cancer	34
Carboplatin	Cancer	20
Paclitaxel	Cancer	26
Etoposide	Inflammatory Response	24

Table 4.7. Top Molecular and Cellular functions derived from a *Core Analysis* performed in Ingenuity Pathway Analysis. The corresponding functions were selected based on the degree of overlap between the miRNA selected in that treatment group and those defined in the IPA database as being associated with that function beyond random chance ($p<0.05$). These results, unlike those represented in Table 4.5, do not factor the degree of connectivity.

Treatment Group	Function	# Molecules
Cisplatin	Cell Death	9
Carboplatin	Cell Cycle	3
Paclitaxel	Cellular Growth and Proliferation	10
Etoposide	Cellular Development	26

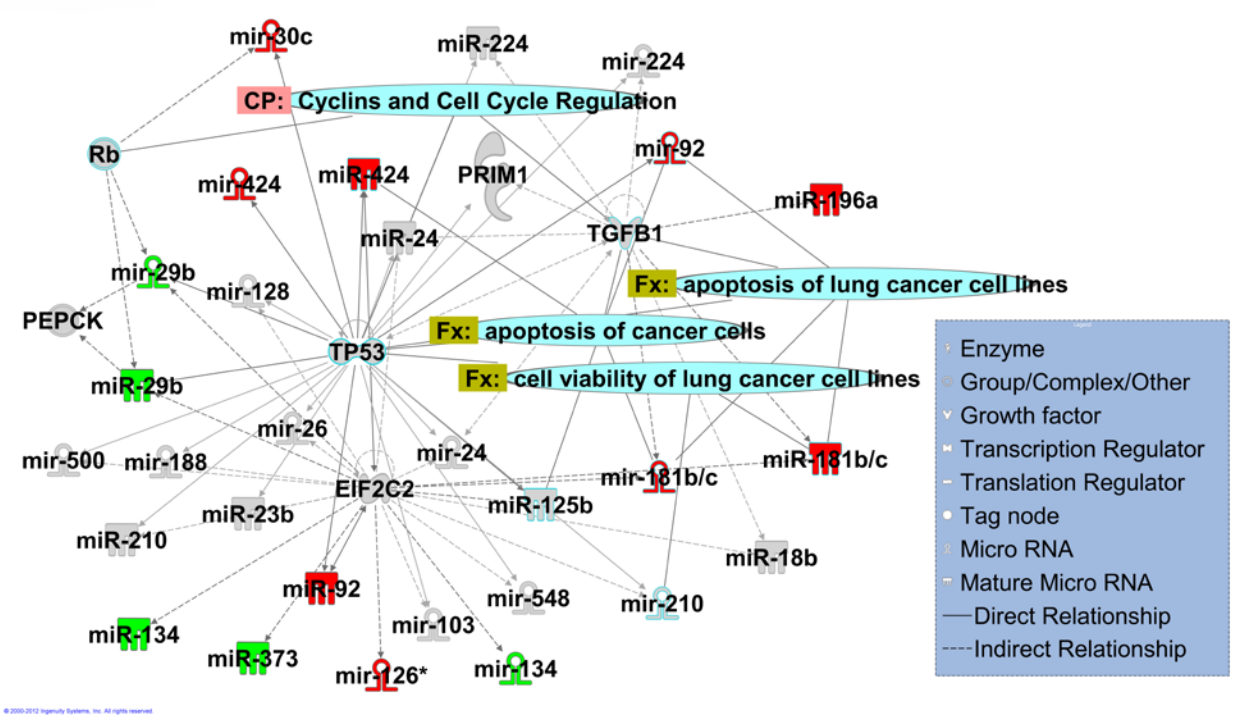

Figure 4.5. Overview of network derived from IPA Core analysis, representing interactions between miRNA found to be chemopredictive for cisplatin and species relevant to the significant functions and diseases listed in Table 4.5. Canonical Pathways (CP:) and Functions and Diseases (Fx:) as defined by IPA are overlaid to highlight the interconnectedness of the set with molecular processes relevant to chemoresponse. Predictive miRNA whose expression was positively associated with survival are shown in green, and those who expression was negatively associated are shown in red.

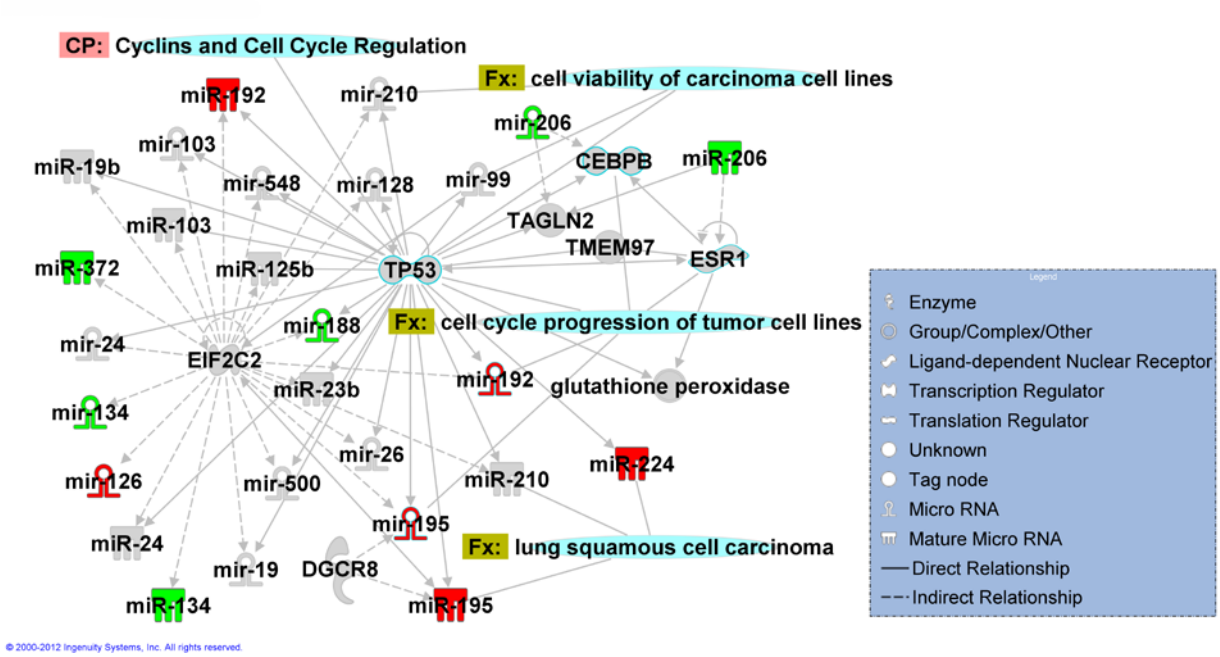

Figure 4.6. Overview of network derived from IPA Core analysis, representing interactions between miRNA found to be chemopredictive for carboplatin and species relevant to the significant functions and diseases listed in Table 4.5. Canonical Pathways (CP:) and Functions and Diseases (Fx:) as defined by IPA are overlaid to highlight the interconnectedness of the set with molecular processes relevant to chemoresponse. Predictive miRNA whose expression was positively associated with survival are shown in green, and those who expression was negatively associated are shown in red.

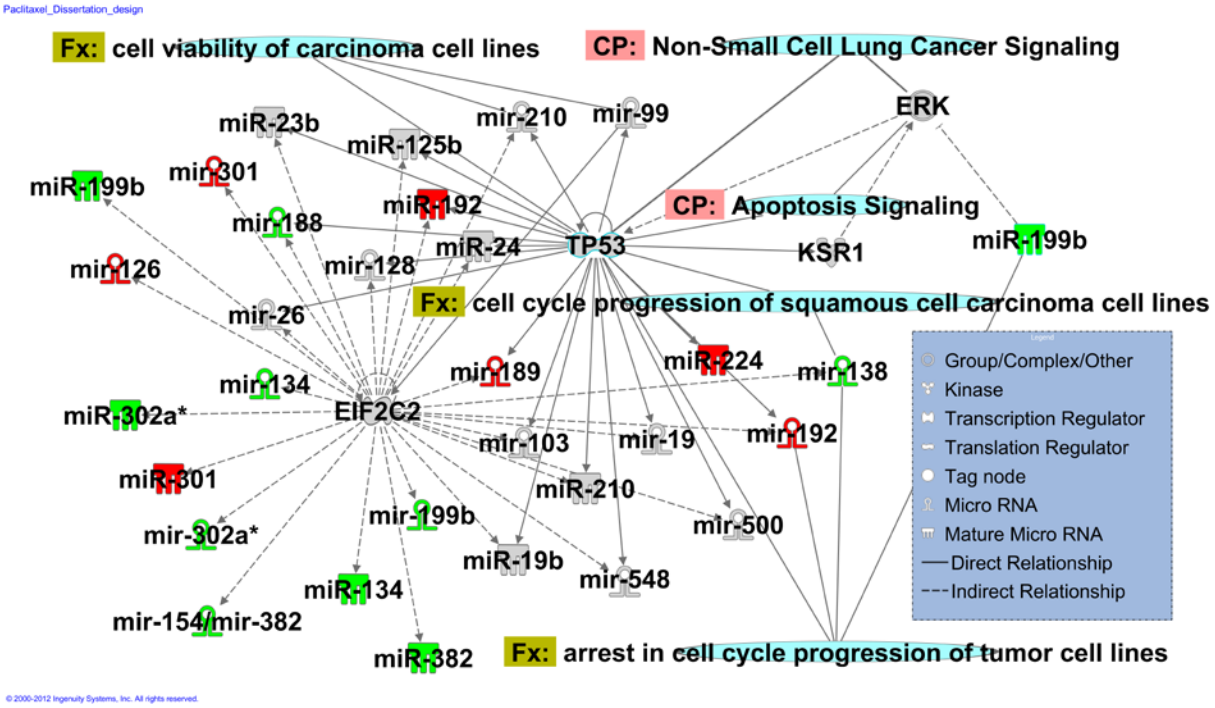

Figure 4.7. Overview of network derived from IPA Core analysis, representing interactions between miRNA found to be chemopredictive for paclitaxel and species relevant to the significant functions and diseases listed in Table 4.5. Canonical Pathways (CP:) and Functions and Diseases (Fx:) as defined by IPA are overlaid to highlight the interconnectedness of the set with molecular processes relevant to chemoresponse. Predictive miRNA whose expression was positively associated with survival are shown in green, and those who expression was negatively associated are shown in red.

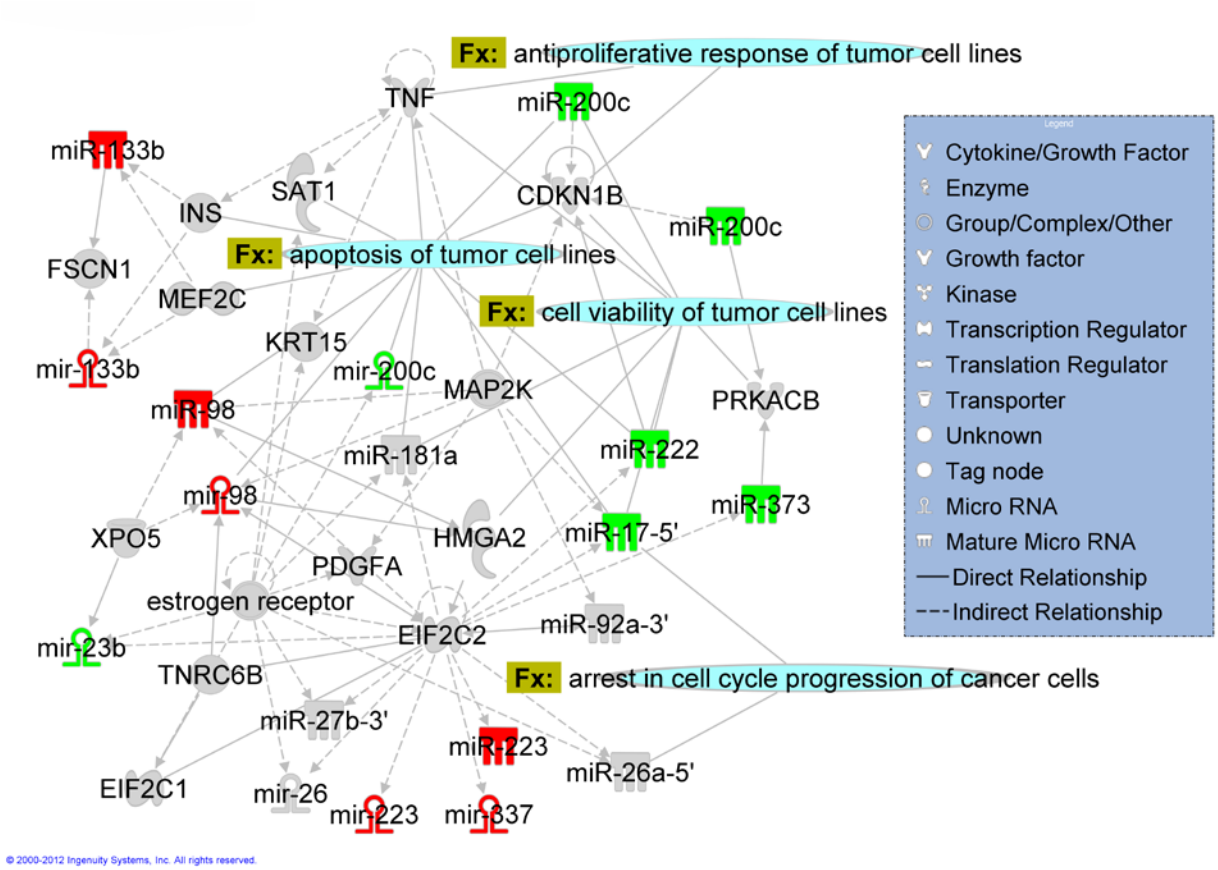

Figure 4.8. Overview of network derived from IPA Core analysis, representing interactions between miRNA found to be chemopredictive for etoposide and species relevant to the significant functions and diseases listed in Table 4.5. Canonical Pathways (CP:) and Functions and Diseases (Fx:) as defined by IPA are overlaid to highlight the interconnectedness of the set with molecular processes relevant to chemoresponse. Predictive miRNA whose expression was positively associated with survival are shown in green, and those who expression was negatively associated are shown in red.

In most cases miRNA had a similar association with survival across the surgery and radiation treatment groups, although some miRNA gained or lost significance depending upon the treatment. This may be due to moderation of the effect by treatment type, but is likely due at least in part to differing sample sizes between treatment modalities. In general, miRNA with highly similar sequences had a similar association with survival, even if one or more of the similar miRNA failed to achieve significance in a given analysis, i.e. the miR-181b/c group and the 3' and 5' strands of miR-17. This may indicate that targets shared by these miRNA may be more central to the observed mediation than those affected by only a single miRNA. Additionally, many of the results from the IPA analyses indicate that these sets of miRNA, and their corresponding networks, play a role in apoptosis, cellular growth, and proliferation. All four of the significant sets from the analyses on individual agents had aspects of cellular growth or death as their top biological function. In addition, an overlay of canonical pathways and cellular functions shows that many of the networked molecules derived from the same sets are also related to apoptosis or cell-cycle regulation (Figures 4.5 to 4.8). This suggests that the significant miRNA may act in a broad pattern of expression to modulate the response of tumor cells to chemotherapeutic agents, as indicated by improved or diminished survival by patients treated with those agents.

There are however multiple limitations to this approach. The first is the assumption that miRNA expression patterns in patients in the Raponi cohort are representative of those in other patients with similar disease status. The accuracy to which these expression patterns can be estimated is also limited by the smaller sample size. Adequate sample size is a problem inherent to similar studies to date. In addition, the Raponi cohort lacks patients with stage 4 tumors. Although the benefit of chemotherapy in Stage 4 cases is better characterized than in earlier stages, the absence of stage 4 data for our method nonetheless affects the generalizability of the results.

The use of administrative data also has some important implications. Although the SEER data are in general highly accurate, it is not fully inclusive of all treatments a patient may receive. Treatments not covered by Medicare or covered by other forms of insurance would not appear in the database. Additionally, eligibility requirements for Medicare coverage artificially limit the patient sample to those over the age of 65, with notable exceptions such as eligibility due to disability benefits. However, as lung cancer occurs primarily in older populations the effect is limited relative to other cancer types which are common in comparatively younger populations.

By design, the SEER registry represents a more demographically and clinically diverse group of patients when compared to cohorts limited to a specific geographic area or healthcare system. Finally, it

should be noted that many of the miRNA found to be predictive or prognostic lacked any significant annotation which would allow for their potential biological roles to be ascertained, so the exclusion of a particular miRNA does not necessarily indicate a null effect.

Despite these limitations, this approach allowed for the identification of miRNA which were useful for both prognostication and prediction of response to chemotherapy through consideration of both clinical and population-based measures of survival, in addition to cellular response and biological context. Stratification of the analyses on treatment modality mitigates the effect of differential treatment assignment on the length of survival. This allows for ascertainment of the contribution of miRNA markers to survival and chemoresponse independent of treatment effects. Further refinement and validation of this approach holds the potential for identification of miRNA regulatory pathways which may mediate chemoresponse and survival in a large number of patients.

Chapter 5:

Conclusion

Two unresolved issues in the treatment of lung cancer, assessment of risk of recurrence and the selection of an effective chemotherapeutic agent for patients with similar tumor morphology, were identified as potential areas for the application of prognostic and predictive models of patient survival. It was shown that given an appropriate theoretical framework, each of the major unresolved issues can be approached through the use of comprehensive prognostic and predictive models incorporating demographic, clinical, and biomarker information.

The second chapter developed an analytical framework for using registry information to develop a prognostic model for lung cancer patients. Information on tumor stage, grade, patient age, race, and sex was used to construct a comprehensive model of patient survival which was shown to be superior to a similar model using tumor stage alone. The model which was developed offered measurable improvements in prognostication across a wide variety of patients and was able to be translated into an intuitive tool suitable for use in clinical practice. It was shown that reliable information on tumor morphology, demographic information, and follow-up can be derived for a larger set of patients than would be feasible to enroll in a prospective study. The larger patient sample allowed for a more robust assessment of internal validity and assessment within treatment modality. The analytical framework established in Chapter 2 also served as a basis for further extending the comprehensive prognostic model in the third chapter.

The third chapter expanded the prognostic ability of the previously defined model by assessing the role of co-morbidity in affecting disease-specific survival and the effect of chemotherapy on prognostication. It was found that a variety of co-morbid conditions can affect survival. Of these, COPD was chosen due to its strong independent effect on survival, improvements in prognostication in a comprehensive model, and persistence of these effects within treatment sub-groups. The results show that COPD has a dual effect both as a confounder for treatment candidacy and as a driver of both long and short-term patient survival. Because of this, COPD was chosen to be added to the previously defined model. Through the use of linked Medicare data, it was also possible to assess the effect of COPD status on survival and the prognostic ability of the previously defined comprehensive model in groups of patients receiving chemotherapy.

The web-based application of the models developed in Chapters 2 and 3 demonstrated how the comprehensive prognostic model can be translated into a useful and intuitive tool for clinicians. The

comprehensive model of patient survival presented In Chapters 2 and 3 incorporated multiple demographic and clinicopathologic variables to create a single measure of patient risk, a Hazard Score, used in estimating expected survival at various timepoints. This single Hazard Score was also used to estimate survival relative to other patients, and as a factor for the selection of high-and low-risk patients assesed using the web-based prognostic model.

The fourth chapter demonstrated that the projection of miRNA expression onto a population cohort from a comparatively smaller group of patients can be a useful approach to validating and selecting miRNA indicative of influencing chemoresponse and prognosis. The large, well-annotated, and diverse population of patients in the linked SEER-Medicare database served as a secondary source of patient survival data. By linking the clinical and population cohorts through the use of shared tumor progression categories it was possible to select a set of miRNA which showed a strong association with survival in patients receiving a variety of chemotherapeutic, surgical, and radiological treatments. These miRNA sets were then further validated using cell-line data, resulting in a set of miRNA which exhibited strong interconnectivity with molecules known to play a role in processes pertinent to response to a chemotherapeutic agent.

Future analyses using this framework would benefit from the refinement of COPD and treatment variables. Examining the role of COPD in biasing the type of surgery performed, i.e. lobe or wedge resection, and the effect of COPD after receiving each of these treatments will give better insight into how the disease influences survival due to treatment candidacy and as a prognostic factor. Similarly, the specific type of radiation administered and the associated outcomes may also be influenced by COPD. New procedures such as Stereotactic Radioablation or Cyber-Knife offer improved outcomes over traditional radiotherapy (Brown et al., 2007a; Brown et al., 2007b), and may greatly benefit patients whose tumor may be inoperable due to COPD. It may also be possible to gauge the severity of COPD in patients identified as having the condition. One proposed method would use oxygen dependence status to split COPD patients into two groups; an oxygen dependent group with presumably more advanced disease, and another group not dependent on oxygen with presumably less advanced disease. This refinement in COPD status paired with determining the specific type of surgery and radiation received may help to further elucidate the role of COPD in influencing treatment candidacy and survival.

Reference List

American Cancer Society (2012). *Cancer Facts & Figures 2012* American Cancer Society.

American Joint Committee on Cancer (2010). *AJCC Cancer Staging Manual*. New York: Springer-Verlag.

Avraham, R. & Yarden, Y. (2012). Regulation of signalling by microRNAs. *Biochemical Society Transactions, 40,* 26-30.

Bach, P. B., Cramer, L. D., Warren, J. L., & Begg, C. B. (1999). Racial differences in the treatment of early-stage lung cancer. *New England Journal of Medicine, 341,* 1198-1205.

Beer, D. G., Kardia, S. L. R., Huang, C. C., Giordano, T. J., Levin, A. M., Misek, D. E. et al. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine, 8,* 816-824.

Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P. et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc.Natl.Acad.Sci.U.S.A, 98,* 13790-13795.

Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D. et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature, 439,* 353-357.

Birim, O., Kappetein, A. P., Waleboer, M., Puvimanasinghe, J. P., Eijkemans, M. J., Steyerberg, E. W. et al. (2006). Long-term survival after non-small cell lung cancer surgery: development and validation of a prognostic model with a preoperative and postoperative mode. *J Thorac.Cardiovasc.Surg., 132,* 491-498.

Borczuk, A. C., Kim, H. K., Yegen, H. A., Friedman, R. A., & Powell, C. A. (2005). Lung adenocarcinoma global profiling identifies type II transforming growth factor-beta receptor as a repressor of invasiveness. *Am.J Respir Crit Care Med., 172,* 729-737.

Boyd, M. R. & Pauli, K. D. (1995). Some Practical Considerations and Applications of the National-Cancer-Institute In-Vitro Anticancer Drug Discovery Screen. *Drug Development Research, 34,* 91-109.

Brown, J. S., Eraut, D., Trask, C., & Davison, A. G. (1996). Age and the treatment of lung cancer. *Thorax, 51,* 564-568.

Brown, W. T., Wu, X. D., Amendola, B., Perman, M., Han, H. K., Fayad, F. et al. (2007a). Treatment of early non-small cell lung cancer, stage IA, by image-guided robotic stereotactic radioablation-CyberKnife. *Cancer Journal, 13,* 87-94.

Brown, W. T., Wu, X. D., Fayad, F., Fowler, J. F., Amendola, B. E., Garcia, S. et al. (2007b). CyberKnife (R) radiosurgery for stage I lung cancer: Results at 36 months. *Clinical Lung Cancer, 8,* 488-492.

Bryant, A. & Cerfolio, R. J. (2007). Differences in epidemiology, histology, and survival between cigarette smokers and never-smokers who develop non-small cell lung cancer. *Chest, 132,* 185-192.

Chang, H. C., Cho, C. Y., & Hung, W. C. (2006). Silencing of the metastasis suppressor RECK by RAS oncogene is mediated by DNA methyltransferase 3b-induced promoter methylation. *Cancer Research, 66,* 8413-8420.

Charlson, M. E., Pompei, P., Ales, K. L., & Mackenzie, C. R. (1987). A New Method of Classifying Prognostic Co-Morbidity in Longitudinal-Studies - Development and Validation. *Journal of Chronic Diseases, 40,* 373-383.

Chen, H. Y., Yu, S. L., Chen, C. H., Chang, G. C., Chen, C. Y., Yuan, A. et al. (2007). A five-gene signature and clinical outcome in non-small-cell lung cancer. *New England Journal of Medicine, 356,* 11-20.

Chen, X., Hu, Z. B., Wang, W. J., Ba, Y., Ma, L. J., Zhang, C. N. et al. (2012). Identification of ten serum microRNAs from a genome-wide serum microRNA expression profile as novel noninvasive biomarkers for nonsmall cell lung cancer diagnosis. *International Journal of Cancer, 130,* 1620-1628.

Chen, Y. L., Chang, M. C., Huang, C. Y., Chiang, Y. C., Lin, H. W., Chen, C. A. et al. (2011). Serous ovarian carcinoma patients with high alpha-folate receptor had reducing survival and cytotoxic chemo-response. *Molecular Oncology*.

Clegg, L. X., Li, F. P., Hankey, B. G., Chu, K., & Edwards, B. K. (2002). Cancer survival among US whites and minorities - A SEER (Surveillance, Epidemiology, and End Results) program population-based study. *Archives of Internal Medicine, 162,* 1985-1993.

Coleman, R., Herzog, T., Hosford, S., Sevin, B. U., Wells, A., Fiori, L. et al. (2004). An ex vivo chemoresponse assay predicts carboplatin-induced progression-free interval in ovarian cancer. *Journal of Clinical Oncology, 22,* 482S.

Dalton, W. S. & Friend, S. H. (2006). Cancer biomarkers--an invitation to the table. *Science, 312,* 1165-1168.

Dehing-Oberije, C., De Ruysscher, D., van der Weide, H., Hochstenbag, M., Bootsma, G., Geraedts, W. et al. (2008). Tumor volume combined with number of positive lymph node stations is a more important prognostic factor than TNM stage for survival of non-small-cell lung cancer patients

treated with (chemo)radiotherapy. *International Journal of Radiation Oncology Biology Physics, 70,* 1039-1044.

Dehing-Oberije, C., Yu, S., De Ruysscher, D., Meersschout, S., Van Beek, K., Lievens, Y. et al. (2009). Development and External Validation of Prognostic Model for 2-Year Survival of NonΓÇôSmall-Cell Lung Cancer Patients Treated With Chemoradiotherapy. *International Journal of Radiation Oncology*Biology*Physics, 74,* 355-362.

Deyo, R. A., Cherkin, D. C., & Ciol, M. A. (1992). Adapting A Clinical Comorbidity Index for Use with Icd-9-Cm Administrative Databases. *Journal of Clinical Epidemiology, 45,* 613-619.

DHoore, W., Bouckaert, A., & Tilquin, C. (1996). Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *Journal of Clinical Epidemiology, 49,* 1429-1433.

Elixhauser, A., Steiner, C., Harris, D. R., & Coffey, R. N. (1998). Comorbidity measures for use with administrative data. *Medical Care, 36,* 8-27.

Engels, E. A. (2008). Inflammation in the development of lung cancer: epidemiological evidence. *Expert Review of Anticancer Therapy, 8,* 605-615.

Fabbri, M., Garzon, R., Cimmino, A., Liu, Z., Zanesi, N., Callegari, E. et al. (2007). MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 15805-15810.

Gallion, H., Christopherson, W. A., Coleman, R. L., Demars, L., Herzog, T., Hosford, S. et al. (2006). Progression-free interval in ovarian cancer and predictive value of an ex vivo chemoresponse assay. *International Journal of Gynecological Cancer, 16,* 194-201.

Gaur, A., Jewell, D. A., Liang, Y., Ridzon, D., Moore, J. H., Chen, C. F. et al. (2007). Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Research, 67,* 2456-2468.

Guo, L., Ma, Y., Ward, R., Castranova, V., Shi, X., & Qian, Y. (2006). Constructing molecular classifiers for the accurate prognosis of lung adenocarcinoma. *Clin.Cancer Res., 12,* 3344-3354.

Guo, N. L., Wan, Y. W., Tosun, K., Lin, H., Msiska, Z., Flynn, D. C. et al. (2008). Confirmation of gene expression-based prediction of survival in non-small cell lung cancer. *Clin.Cancer Res., 14,* 8213-8220.

Hoffman, P. C., Mauer, A. M., & Vokes, E. E. (2000). Lung cancer. *Lancet, 355,* 479-485.

Horio, Y., Takahashi, T., Kuroishi, T., Hibi, K., Suyama, M., Niimi, T. et al. (1993). Prognostic significance of p53 mutations and 3p deletions in primary resected non-small cell lung cancer. *Cancer Res., 53,* 1-4.

Houghton, A. M., Mouded, M., & Shapiro, S. D. (2008). Common origins of lung cancer and COPD. *Nature Medicine, 14,* 1023-1024.

Iorio, M. V. & Croce, C. M. (2012). microRNA involvement in human cancer. *EMBO Molecular Medicine, 4,* 143-159.

Janssen-Heijnen, M. L. G., Schipper, R. M., Razenberg, P. P. A., Crommelin, M. A., & Coebergh, J. W. W. (1998). Prevalence of co-morbidity in lung cancer patients and its relationship with treatment: A population-based study. *Lung Cancer, 21,* 105-113.

Janssen-Heijnen, M. L. G., Smulders, S., Lemmens, V. E. P. P., Smeenk, F. W. J. M., van Geffen, H. J. A. A., & Coebergh, J. W. W. (2004). Effect of comorbidity on the treatment and prognosis of elderly patients with non-small cell lung cancer. *Thorax, 59,* 602-607.

Jemal, A., Siegel, R., Ward, E., Hao, Y. P., Xu, J. Q., Murray, T. et al. (2008). Cancer statistics, 2008. *Ca-A Cancer Journal for Clinicians, 58,* 71-96.

Jung, M., Schaefer, A., Steiner, I., Kempkensteffen, C., Stephan, C., Erbersdobler, A. et al. (2010). Robust MicroRNA Stability in Degraded RNA Preparations from Human Tissue and Cell Samples. *Clinical Chemistry, 56,* 998-1006.

Kenfield, S., Wei, E., Colditz, G., Stampfer, M., & Rosner, B. (2006). Comparison of aspects of smoking among four histologic types of lung cancer. *American Journal of Epidemiology, 163,* S108.

Kiri, V. A., Soriano, J., Visick, G., & Fabbri, L. (2010). Recent trends in lung cancer and its association with COPD: an analysis using the UK GP Research Database. *Primary Care Respiratory Journal, 19,* 57-61.

Klabunde, C. N., Potosky, A. L., Legler, J. M., & Warren, J. L. (2000). Development of a comorbidity index using physician claims data. *Journal of Clinical Epidemiology, 53,* 1258-1267.

Kodama, K., Ishii, G., Miyamoto, S., Goya, M., Zhang, S. C., Sangai, T. et al. (2005). Laminin 5 expression protects against anoikis at aerogenous spread and lepidic growth of human lung adenocarcinoma. *International Journal of Cancer, 116,* 876-884.

Liu, Y. Y., Xiao, H., Tian, Y. M., Nekrasova, T., Hao, X. P., Lee, H. J. et al. (2008). The Pak4 protein kinase plays a key role in cell survival and tumorigenesis in athymic mice. *Molecular Cancer Research, 6,* 1215-1224.

Ludwig, J. A. & Weinstein, J. N. (2005). Biomarkers in cancer staging, prognosis and treatment selection. *Nat.Rev.Cancer, 5,* 845-856.

Mak, G. M., Chan, M. M. L., Leong, V. Y. L., Lee, J. M. F., Yau, T. O., Ng, I. O. L. et al. (2011). Overexpression of a Novel Activator of PAK4, the CDK5 Kinase-Associated Protein CDK5RAP3, Promotes Hepatocellular Carcinoma Metastasis. *Cancer Research, 71,* 2949-2958.

Manda, R., Kohno, T., Niki, T., Yamada, T., Takenoshita, S., Kuwano, H. et al. (2000). Differential expression of the LAMB3 and LAMC2 genes between small cell and non-small cell lung carcinomas. *Biochemical and Biophysical Research Communications, 275,* 440-445.

Markou, A., Tsaroucha, E. G., Kaklamanis, L., Fotinou, M., Georgoulias, V., & Lianidou, E. S. (2008). Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. *Clinical Chemistry, 54,* 1696-1704.

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L. et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 10513-10518.

Mraz, M., Malinova, K., Mayer, J., & Pospisilova, S. (2009). MicroRNA isolation and stability in stored RNA samples. *Biochemical and Biophysical Research Communications, 390,* 1-4.

Naruke, T., Goya, T., Tsuchiya, R., & Suemasu, K. (1989). Prognosis and survival in resected lung carcinoma based on the new international staging system. *J Thorac Cardiovasc Surg*.

Olivotto, I. A., Bajdik, C. D., Ravdin, P. M., Speers, C. H., Coldman, A. J., Norris, B. D. et al. (2005). Population-based validation of the prognostic model ADJUVANT! for early breast cancer. *J Clin.Oncol., 23,* 2716-2725.

Orourke, M. A., Feussner, J. R., Feigl, P., & Laszlo, J. (1987). Age Trends of Lung-Cancer Stage at Diagnosis - Implications for Lung-Cancer Screening in the Elderly. *Jama-Journal of the American Medical Association, 258,* 921-926.

Papi, A., Casoni, G., Caramori, G., Guzzinati, I., Boschetto, P., Ravenna, F. et al. (2004). COPD increases the risk of squamous histological subtype in smokers who develop non-small cell lung carcinoma. *Thorax, 59,* 679-681.

Pepe, M. S., Janes, H., Longton, G., Leisenring, W., & Newcomb, P. (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology, 159,* 882-890.

Potosky, A. L., Riley, G. F., Lubitz, J. D., Mentnech, R. M., & Kessler, L. G. (1993). Potential for Cancer-Related Health-Services Research Using A Linked Medicare-Tumor Registry Database. *Medical Care, 31,* 732-748.

Putila, J., Remick, S. C., & Guo, N. L. (2011). Combining Clinical, Pathological, and Demographic Factors Refines Prognosis of Lung Cancer: A Population-Based Study. *Plos One, 6*.

Rabinowits, G., Gercel-Taylor, C., Day, J. M., Taylor, D. D., & Kloecker, G. H. (2009). Exosomal MicroRNA: A Diagnostic Marker for Lung Cancer. *Clinical Lung Cancer, 10,* 42-46.

Rancati, T., Ceresoli, G. L., Gagliardi, G., Schipani, S., & Cattaneo, G. M. (2003). Factors predicting radiation pneumonitis in lung cancer patients: a retrospective study. *Radiotherapy and Oncology, 67,* 275-283.

Raponi, M., Zhang, Y., Yu, J., Chen, G., Lee, G., Taylor, J. M. G. et al. (2006). Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Research, 66,* 7466-7472.

Raponi, M., Dossey, L., Jatkoe, T., Wu, X., Chen, G., Fan, H. et al. (2009). MicroRNA Classifiers for Predicting Prognosis of Squamous Cell Lung Cancer. *Cancer Research, 69,* 5776-5783.

Ravdin, P. M., Siminoff, L. A., Davis, G. J., Mercer, M. B., Hewlett, J., Gerson, N. et al. (2001). Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *Journal of Clinical Oncology, 19,* 980-991.

Rodenhuis, S., van de Wetering, M. L., Mooi, W. J., Evers, S. G., van Zandwijk, N., & Bos, J. L. (1987). Mutational Activation of the K-ras Oncogene. *New England Journal of Medicine*.

Romano, P. S., Roos, L. L., & Jollis, J. G. (1993). Adapting A Clinical Comorbidity Index for Use with Icd-9-Cm Administrative Data - Differing Perspectives. *Journal of Clinical Epidemiology, 46,* 1075-1079.

Schumacher, M., Graf, E., & Gerds, T. (2003). How to assess prognostic models for survival data: A case study in oncology. *Methods of Information in Medicine, 42,* 564-571.

SEER. (2010). Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: Incidence - SEER 9 Regs Research Data, Nov 2009 Sub (1973-2007) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969-2007 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2010,

based on the November 2009 submission.

Ref Type: Generic

Seiffert, J. (1993). *SEER Program: Comparative staging guide for cancer, version 1.1* (Rep. No. 93-3640). National Cancer Institute.

Shedden, K., Taylor, J. M. G., Enkemann, S. A., Tsao, M. S., Yeatman, T. J., Gerald, W. L. et al. (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature Medicine, 14,* 822-827.

Shoemaker, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer, 6,* 813-823.

Slebos, R. J. C., Kibbelaar, R. E., Dalesio, O., Kooistra, A., Stam, J., Meijer, C. J. L. M. et al. (1990). K-ras Oncogene Activation as a Prognostic Marker in Adenocarcinoma of the Lung. *New England Journal of Medicine*.

Stein, W. D., Litman, T., Fojo, T., & Bates, S. E. (2004). Serial Analysis of Gene Expression (SAGE) database analysis of chemosensitivity: Comparing solid tumors with cell lines and comparing solid tumors from different tissue origins. *Cancer Research, 64,* 2805-2816.

Steyerberg, E. W., Harrell, F. E., Borsboom, G. J. J. M., Eijkemans, M. J. C., Vergouwe, Y., & Habbema, J. D. F. (2001). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology, 54,* 774-781.

Tammemagi, C. M., Neslund-Dudas, C., Simoff, M., & Kvale, P. (2004). Smoking and lung cancer survival - The role of comorbidity and treatment. *Chest, 125,* 27-37.

Thatcher, N., Chang, A., Parikh, P., Rodrigues, P. J., Ciuleanu, T., von, P. J. et al. (2005). Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet, 366,* 1527-1537.

Thorpe, K. E. & Howard, D. (2003). Health insurance and spending among cancer patients. *Health Affairs, 22,* W189-W198.

Visbal, A. L., Williams, B. A., Nichols, F. C., Marks, R. S., Jett, J. R., Aubry, M. C. et al. (2004). Gender differences in non-small-cell lung cancer survival: An analysis of 4,618 patients diagnosed between 1997 and 2002. *Annals of Thoracic Surgery, 78,* 209-215.

Vogt, U., Striehn, E., Bosse, U., Klinke, F., & Falkiewicz, B. (1999). Lack of squamous cell lung carcinoma in vitro chemosensitivity to various drug regimens in the adenosine triphosphate cell viability chemosensitivity assay. *Acta Biochimica Polonica, 46,* 299-302.

Vogt, U., Zaczek, A., Klinke, F., Granetzny, A., Bielawski, K., & Falkiewicz, B. (2002). p53 Gene status in relation to ex vivo chemosensitivity of non-small cell lung cancer. *Journal of Cancer Research and Clinical Oncology, 128,* 141-147.

Wan, Y. W., Beer, D. G., & Guo, N. L. (2012). Signaling pathway-based identification of extensive prognostic gene signatures for lung adenocarcinoma. *Lung Cancer, 76,* 98-105.

Wang, Q. Z., Xu, W., Habib, N., & Xu, R. A. (2009). Potential Uses of MicroRNA in Lung Cancer Diagnosis, Prognosis, and Therapy. *Current Cancer Drug Targets, 9,* 572-594.

Warren, J. L., Harlan, L. C., Fahey, A., Virnig, B. A., Freeman, J. L., Klabunde, C. N. et al. (2002a). Utility of the SEER-Medicare data to identify chemotherapy use. *Medical Care, 40,* 55-61.

Warren, J. L., Klabunde, C. N., Schrag, D., Bach, P. B., & Riley, G. F. (2002b). Overview of the SEER-Medicare data - Content, research applications, and generalizability to the United States elderly population. *Medical Care, 40,* 3-18.

Weinstein, J. N. (2006). Spotlight on molecular profiling: "Integromic" analysis of the NCI-60 cancer cell lines. *Molecular Cancer Therapeutics, 5,* 2601-2605.

Xi, Y. G., Nakajima, G., Gavin, E., Morris, C. G., Kudo, K., Hayashi, K. H. et al. (2007). Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *Rna-A Publication of the Rna Society, 13,* 1668-1674.

Yan, L. X., Wu, Q. N., Zhang, Y., Li, Y. Y., Liao, D. Z., Hou, J. H. et al. (2011). Knockdown of miR-21 in human breast cancer cell lines inhibits proliferation, in vitro migration and in vivo tumor growth. *Breast Cancer Research, 13,* R2.

Yao, H. W. & Rahman, I. (2009). Current concepts on the role of inflammation in COPD and lung cancer. *Current Opinion in Pharmacology, 9,* 375-383.

Young, R. P., Hophins, R. J., Hay, B. A., Epton, M. J., Black, P. N., & Gamble, G. D. (2008). Lung cancer gene associated with COPD: triple whammy or possible confounding effect? *European Respiratory Journal, 32,* 1158-1164.

Young, R. P., Hopkins, R. J., Christmas, T., Black, P. N., Metcalf, P., & Gamble, G. D. (2009). COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. *European Respiratory Journal, 34,* 380-386.

Yu, S. L., Chen, H. Y., Chang, G. C., Chen, C. Y., Chen, H. W., Singh, S. et al. (2008). MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell, 13,* 48-57.