Published in final edited form as:

Clin Lung Cancer. 2013 March ; 14(2): 128–138. doi:10.1016/j.clcc.2012.06.001.

Survival Analysis of Patients With Stage I Non–Small-Cell Lung Cancer Using Clinical and DNA Repair Pathway Expression Variables

Madhusmita Behera¹, John J. Heine², Gabriel L. Sica³, Erin E. Fowler², Ha Tran¹, Robert W. Fu¹, Anthony A. Gal³, Robert Hermann⁴, William Mayfield⁴, Fadlo R. Khuri¹, Taofeek K. Owonikoko¹, and Suresh S. Ramalingam¹

¹Department of Hematology and Medical Oncology, Winship Cancer Institute of Emory University, Atlanta, GA

²Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL

³Department of Pathology and Laboratory Medicine, Winship Cancer Institute of Emory University, Atlanta, GA

⁴WellStar Kennestone Hospital, Marietta, GA

Abstract

Clinical, pathologic, and DNA expression variables for patients with stage I non–small-cell lung cancer (NSCLC) (N = 161) were investigated to determine factors related to favorable survival. Survival and logistic regression (LR) modeling showed that female sex, younger age, and adenocarcinoma (AC) histologic subtype confer better survival. Poly(ADP-ribose) polymerase (PARP) and Ku86 expression was not correlated with prognosis.

Background—Lung cancer is the leading cause of cancer-related mortality. Understanding patient attributes that enhance survival and predict recurrence is necessary to individualize treatment options.

Methods—Patients (N = 162) were dichotomized into favorable (n = 101) and unfavorable (n = 61) groups based on survival characteristics. Ku86 and poly(ADP-ribose) polymerase (PARP) expression measures were incorporated into the analyses. LR, Kaplan-Meier analysis, and Cox regression were used to investigate intervariable relationships and survival. Odds ratios (ORs) and hazard ratios (HRs) with 95% confidence intervals (CIs) were used to assess associations.

Results—Sex (OR, 0.32; CI-0.14, 0.76), squamous cell carcinoma (SCC) (OR, 0.41; CI-0.17, 0.98), and recurrence (OR, 0.04; CI-0.01, 0.20) confer an unfavorable outcome with area under the

---

Address for correspondence: Madhusmita Behera, PhD, Department of Hematology-Oncology, Winship Cancer Institute of Emory University, 1365 Clifton Rd., Atlanta, GA 30322, Fax: 404-778-5520; mbehera@emory.edu.

Disclosure

The authors have stated that they have no conflicts of interest.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.clcc.2011.03.001>://dx.doi.org/10.1016/j.clcc.2011.03.001.

receiver operating characteristic curve (A_z) = 0.788. Patients with increased tumor grade (OR = 1.84; CI-1.06, 3.19) or increased Ku86 intensity (OR, 2.03; CI-1.08, 3.82) were more likely to be male individuals, and older patients (OR, 1.70; CI-(1.14, 2.52) were more likely to have SCC. Patients older than the median age (HR, 1.86; CI-1.11, 3.12), patients with SCC (HR, 1.78; CI-1.05, 3.01), patients with recurrence (HR, 4.16; CI-2.37, 7.31), and male patients (HR, 2.03; CI-1.20, 3.43) have a higher hazard. None of the DNA repair measures were associated with significant HRs.

Conclusion—Clinical and pathologic factors that enhance and limit survival for patients with stage I NSCLC were quantified. The DNA repair measures showed little association. These findings are important given that certain clinical and pathologic features are related to better long-term survival outcome than others.

Keywords

Cox regression; DNA repair; Kaplan-Meier; Ku86; Logistic regression; NSCLC; PARP

Introduction

Lung cancer is the leading cause of cancer-related mortality.¹ Advances in therapeutic modalities have resulted in modest survival. However, a cure remains elusive for patients with disease beyond stage I, which is often the case.^{2,3} Even in patients with early-stage lung cancer, there is a critical need to improve cure rates and identify patients at higher risk for recurrence.

Previous work^{1,4} showed that early stage at diagnosis, younger age, and female sex are favorable prognostic indicators for non-small-cell lung cancer (NSCLC). It is also important to note that the incidence rates for the various forms of lung cancer appear to be shifting in time,⁵ and that there are both racial and regional differences throughout the United States.⁶ Both serial and geographic variations in lung cancer survival patterns indicate that survival rates require continual evaluation to ensure that the knowledge base is current.

In addition to clinical factors, the molecular characteristics of the tumors can be used to determine prognosis. In particular, *repair cross complementing gene 1 (ERCC1)* expression is a prognostic factor in patients with early-stage NSCLC.⁷ *ERCC1* plays an important role in the nucleotide excision repair pathway. Given that DNA repair is mediated by a number of other important pathways as well, we investigated the impact of poly(ADP-ribose) polymerase (PARP) and Ku86 expression in specimens of patients with early-stage NSCLC, along with clinical factors,^{1,4} to identify the variables that either confer or limit survival.^{8,9} We hypothesized that the expression of these proteins could influence prognosis, treatment selection, or possibly recurrence for patients with NSCLC.

Patients and Methods

We analyzed clinical data and selected DNA repair proteins from patients with stage I NSCLC who underwent surgical resection. We used 2 forms of analysis to evaluate the survival characteristics of this population. Logistic regression (LR) was used to study 2

groups of patients dichotomized by their survival characteristics to form favorable and unfavorable survival outcome groups. Intervariable relationships were also explored with LR. Kaplan-Meier and Cox regression survival analyses were used to study various patient strata.

Time-to-event and LR modeling convey different information. Cox regression is not typically used to make estimates at the patient level but can provide an instantaneous relative risk given a set of covariates for a given patient. Developing methods derived from Cox regression for individual estimates is an active field of research.¹⁰ Kaplan-Meier analysis is nonparametric and is not usually applied at the patient level. Reducing the data resolution to a binary outcome makes the data set amenable to both LR modelling, for patient estimates specifically, and to all forms of binary classification applications more generally, which is an alternative approach to survival analysis¹¹ that can serve as a simplifying mechanism.¹² The LR model provides the probability of a predefined endpoint given a set of specific covariates and thus gives an output that is easily interpretable at the patient level.

Study Population and Measures

The data set is composed of patients (N = 162) with stage I NSCLC who underwent resection at the WellStar Kennestone Hospital in Marietta, GA from 2002 to 2008. Patients were selected post hoc and consecutively. Clinical and pathologic data were abstracted from the patient records. The selection criteria included all patients with stage I disease and complete case ascertainment for the variables under consideration. One hundred one (n_1) of these patients were alive at last contact (censored group), and 61 (n_2) patients had died (incident group) during the course of the contact interval. Abstracted variables included age (ie, age of the patient at the time of surgery) with integer accuracy, sex (binary), smoking status (binary), 4 histologic subtypes (adenocarcinoma [AC], squamous cell carcinoma [SCC], large cell carcinoma [LCC], and adenosquamous carcinoma (ASC), tumor grade, adjuvant treatment, and disease recurrence. Smoking history was categorized as either smoker, past or present (yes), or as never smoked (no). Tumor-grade is a 1 to 3 integer scale describing the cancer cell differentiation. Adjuvant treatment defines systemic chemotherapy after surgery. Recurrence defines relapse after surgery.

Experimental DNA Expression Biomarker Measures

DNA repair protein expression for Ku86 and PARP were evaluated as disease biomarkers from digitally scanned, immunohistochemically stained NSCLC tissue microarrays (TMAs). Before the creation of the TMAs used for this study, a board-certified pathologist (GS) reviewed selected hematoxylin and eosin (H&E)-stained slides to confirm the diagnosis and grade of the tumors. Areas of interest were identified on the H&E-stained slides, and tissue cores were obtained from the corresponding areas of the originating formalin-fixed paraffin-embedded tissue block using a semiautomated tissue microarrayer (Pathology Devices, Inc, Westminster, MD). Because of histologic heterogeneity, triplicate tumor tissue cores were obtained from each patient. The 4- μ m-thick sections from the TMA blocks were stained with the Ku86 (sc-56136, Santa Cruz Biotechnology, Inc, Santa Cruz, CA) and PARP (Clontech, Mountain View, CA) monoclonal antibodies according to manufacturer's instructions using an automated stainer (Dako, Carpinteria, CA). The stained TMA slides

were digitized with a Nanozoomer whole slide scanner (Olympus America, Inc, Center Valley, PA). Stained TMA pathologic images were scored using a modified scoring methodology in which percentage staining and intensity was visually assessed by the pathologist for each core. A final intensity score for each patient was determined by averaging the 3 scores. This method of scoring was repeated for every patient. The 3 measures were derived for each DNA repair protein expression that measured intensity (I), proportion (P), and total score (S) (ie, intensity $\times$ proportion). For Ku86, we refer to these as K_I , K_P , and K_S , respectively. Similarly for PARP, we refer to these as P_I , P_P , and P_S , respectively.

Analysis Methods

Favorable Outcome Analysis—Two-class predictive (or prognostic) models were developed with LR. Censored (n_1) and incident (n_2) patients were used to form favorable and unfavorable survival outcome groups, respectively, as described and validated previously. The LR model was referenced to the favorable outcome group (ie, to predict the probability of a given patient experiencing a favorable survival outcome). Complete case ascertainment for all the variables for the entire patient population was not available. We studied the full data set (full group) and subgroups of this data set, depending on the case ascertainment for the variables under investigation. The goal was to find those variables that related to favorable survival outcome and characterize their association strengths. Odds ratios (ORs) were used to assess associations and the area under the receiver operating characteristic curve (Az) was used to measure predictive capability. The ORs are cited with 95% confidence intervals (CIs).

Intervariable Association Analysis—Various combinations of variables were used to evaluate possible associations with the AC and SCC histologic subtype, sex, and disease recurrence. In this modeling, histologic sub type, sex, and recurrence were used as the dependent variables for LR. We investigated LR models with the following variables: age (A), tumor grade (Gr), sex (S), and the DNA expression measures. Each of these clinical variables are generically referred to as CL1, and each of the 6 DNA repair protein expression variables are generically referred to as PR in this description. We investigated the following relationships to predict the 2 class histologic subtypes (ie, predict SCC): LR(A, Gr), LR(A, S), LR(CL1, PR), LR(A, Gr, S), LR(A, Gr, PR), LR(A,S,PR), and LR (A, Gr,S,PR). This evaluation included 56 different LR models. We performed a similar analysis to predict male sex. The dependent variables used in the LR model to predict male sex can be determined from the histologic (H) subtype analysis by replacing S with H (ie, 56 different LR models).

We investigated various relationships to predict recurrence. We refer to A, Gr, H, and S individually as CL2 clinical variables in this description. We investigated, LR(CL2), LR(PR), LR(CL2, PR), LR(A, Gr), LR(A, H), LR(A, S), LR(CL2, PR), LR(A, Gr, H), LR(A, Gr, S), LR(A, H, S), LR(A, Gr, PR), LR(A, H, PR), LR(A, S, PR), LR(Gr, H, PR), LR(Gr, S, PR), LR(H, S, PR), LR(A, Gr, H, PR), LR(A, H, S, PR), LR(Gr, H, S, PR), and LR(A, Gr, H, S, PR). This evaluation included 110 LR models.

Survival Analysis—Kaplan-Meier survival probability analysis was applied to evaluate survival differences between various patient strata. Hazard ratios (HRs) were used to assess group survival characteristics with 95% CIs using Cox regression analysis. To study age-related survival, the patients were dichotomized using the population median age as the cutoff point and using the below-median age group as the reference. We also dichotomized the patient population by 2-group histologic subtype (ie, SCC and AC), recurrence, adjuvant treatment, and sex; the respective references were AC, no recurrence, no adjuvant treatment, and female sex. The 6 DNA repair protein expression measures were also evaluated by dichotomizing the patient population by the respective median value (low and high groups) for each measure using the below-median groups as the reference. To assess possible cosegregation, we stratified by considering those patients with above-median values for 2 given DNA measures simultaneously compared with the remaining population with the same 2 measures. The patient sample numbers in the upper and lower median strata (n_{upper} , n_{lower}) for each comparison were 60,102 for K_P - P_P , 66,106 for K_S - P_S , 66,106 for K_P - P_S , and 60,102 for K_S - P_P . We also restricted the DNA survival analysis to those patients with SCC ($n_{SCC} = 48$) and AC ($n_{AC} = 93$) separately by dichotomizing the respective histologic subgroup at the median value of each protein expression measure (below median as the reference). Additionally, we investigated stage I subgroups in various strata. Patients were stratified by (1) IA and IB using IA as the reference using the full group data set, (2) lower age group patients with stage IA histologic sub type as the reference compared with the remaining patients in the full group data set, (3) all patients with stage IA disease with AC as the reference compared with those remaining patients in the full group, and (4) lower age group patients with both stage IA disease and AC as the reference compared with the remaining patients in the full group (fewer than 4 patients with ASC and 4 patients with unknown histologic type).

Analyses were performed in the SAS programming environment (SAS Institute, Cary, NC).

Results

Patient Characteristics

The patient characteristics are summarized in Table 1 as totals and by incident and censored groups. The median age was 67 years. Censored patients were younger and more likely to have grade 1 tumors, whereas the other grades were similar across groups. There was a near equal representation of men and women. The censored patients were more likely women, who were more likely to have AC rather than SCC histologic sub type. Although only a small number of patients had either ASC or LCC histologic subtypes, patients in the censored group were more likely to have LCC. Approximately 20% of the patients were nonsmokers and this status was similar across groups. Patients in the incident group were more likely to have experienced disease recurrence (qualified further on) and received adjuvant treatment. Tumor location was similar across groups, and censored patients were more likely to have stage IA disease. The Ku86 and PARP expression for each of the 3 measures was similar across the 2 groups. The K_I findings were excluded because most patients had $K_I = 3$. The mean censored time for the favorable group was 3.94 years and the

overall survival time for the unfavorable group was 2.15 years (not shown), indicating the validity of the stratification.

The recurrence variable requires further qualification. Although there are more patients with recurrence in the incident group, relatively few patients in the total population have recurrence. The minimum and maximum known recurrence times were 20 and 1352 days (not shown). Because the recurrence status is unknown for the favorable group past their censored time, the recurrence outcome is within approximately 3.5 years, which is valid when considering the mean censoring time.

Favorable Outcome Analysis

For the full group data set ($N = 162$, with $n_1 = 101$ and $n_2 = 61$), we had complete case ascertainment for age, sex, adjuvant treatment, tumor grade, stage I subgroup, and DNA repair proteins. The forward stepwise selection procedure resulted in a bivariate model. As shown in Table 2, the ORs for age and sex were significant (ie, the CIs do not include unity). When adjusting for sex, the age association (OR, 0.64 per standard deviation [SD] increase) and sex association (OR, 0.39) show that increasing age and male sex confer unfavorable survival outcome (ie, female patients are 2.6 times more likely to be in the favorable group, and younger patients are 1.5 times more likely to be in the favorable group). In this age-sex model, A_z was 0.683. The other measures were not significant independent predictive factors (ie, $A_z < 0.600$) and their OR associations were not significant.

In the subgroup 1 dataset ($n = 149$, with $n_1 = 91$ and $n_2 = 58$), we had complete ascertainment for age, sex, tumor grade, tumor location, histologic subtype, and DNA repair protein expression measures, as shown in Table 3. The univariate analysis found significant associations for age (OR, 0.65 per SD increase] and sex (OR, 0.32 per unit increase], indicating that increasing age and male sex confer an unfavorable outcome. The forward selection process resulted in a bivariate model with age and sex, which had similar ORs as the univariate models, and $A_z = 0.680$. None of the measures were significant.

The subgroup 2 data set ($n = 134$ with $n_1 = 80$ and $n_2 = 54$) was composed of patients with full case ascertainment for the SCC and AC histologic subtypes in conjunction with age, sex, tumor location in the lung, adjuvant treatment, tumor grade, stage I subgroup, and DNA repair proteins expression ratings. The findings are shown in Table 4. The univariate analysis found significant findings for sex (OR, 0.34; $A_z = 0.630$) and histologic subtype (OR, 0.44; $A_z = 0.594$). The forward stepwise procedure resulted in a bivariate sex and histologic subtype model. In this model we found the following: (1) the ORs were similar to those of the respective univariate models, (2) the combined A_z increased to 0.668, and (3) the OR for Histologic sub type was not significant. The findings show that both AC and female sex confer favorable survival outcome, and these 2 variables in combination provided an increased A_z in comparison with either in isolation. The other variables were not significant (Table 4). It is important to note that age was not included in the selection process. The reasons are investigated in the following paragraph.

Different combinations of variables were investigated to determine models with increased predictive capability using the forward stepwise procedure. We refer to this as the best

model data set ($n = 123$, with $n_1 = 71$ and $n_2 = 52$). This evaluation included patients with complete case ascertainment for age, sex, adjuvant treatment, tumor grade, stage subgroup, SCC and AC histologic subtypes, tumor location, DNA repair protein expression measures, smoking status, and recurrence. The findings are shown in Table 5. The model with the greatest predictive capability included sex, histologic sub type, and recurrence, which resulted in a combined Az of 0.788. In this model, the associations for sex (OR, 0.32), histologic subtype (OR, 0.40), and recurrence (OR, 0.04) were significant and all conferred unfavorable outcome. The respective ORs do not vary much from their respective univariate values, indicating that they provide *independent* contributions. When forcing age into the selected model, its association was not significant, but it increased the model's predictive capability, with Az = 0.796 (last row in Table 5).

Interaction Analysis

To understand possible interactions between age and DNA repair protein expression we performed intermeasurement analyses using the subgroup 2 patients. This analysis was portioned into 3 outcomes by considering those variables that could predict the following: (1) the 2-subgroup Histologic sub type [ie, AC or SCC], (2) sex, and (3) recurrence. Only models that were associated with Az = 0.600 are shown when the corresponding ORs were not significant.

The histologic relationships are shown in Supplemental Table 1. Although there are many models that provided some predictive capability, the majority of the OR associations were not significant. In all models that contained age, increasing age was associated with SCC. The association for age in isolation was associated (OR, 1.70 per SD increase) with Az = 0.637. The associations for the PARP and Ku86 expression measures were not significant, although the sex association (OR, 2.17) gained significance in the LR(Gr, S, K_S) model, indicating men are more likely to have the SCC histologic subtype than are women.

The sex associations are shown in Supplemental Table 2. The majority of these associations were not significant, except those provided by grade and Ku86 expression. Although grade in isolation provided weak predictive capability (Az = 0.596), its association (OR, 1.84 per unit increase) indicates that increasing grade is significantly related to male sex; in models that included grade, similar associations were found (ie, 1.46–2.01 range of ORs). Similarly, K_S in isolation provided a significant association (OR, 2.03 per SD increase) with Az = 0.635, indicating a relationship with sex (ie, increasing K_S is related to male sex). In most models that included K_S, it provided significant associations (ie, 1.84–2.14 range of ORs). Histologic subtype gained significance when including K_S. In this bivariate model, the associations for histologic subtype (OR, 2.14) and K_S (OR, 2.05 per SD increase) with Az = 0.681 show that male patients are more likely to have SCC and an increased K_S measure. To understand interactions with disease recurrence, we investigated 110 LR models and found only weak predictors of recurrence, and no significant OR relationships were noted (not shown in a table because ORs were not significant). In summary, none of the univariate models had an Az > 0.582 (provided by tumor grade), none of the bivariate models had an Az > 0.619 (grade and P_I), none of the trivariate models had an Az > 0.625 (grade, sex, P_I), and none of the models with 4 variates had an Az greater than the best trivariate model.

Survival Analysis

Figure 1 shows the dichotomous age grouping survival curves. The findings are shown in Table 6. The upper age group is at significantly greater hazard (HR, 1.86). This table also provides proportional estimates for patients surviving past 3, 5, and 7 years. This shows that 64% of the lower age group survived past 5 years, whereas 47% of the upper age group survived past this time. Controlling for grade in the age hazard caused an insignificant change (HR, 1.93; CI, 1.11, 3.12). Figure 2 shows the survival curves for patients with SCC Histologic sub type compared with patients with AC Histologic sub type. Patients with SCC are at a significantly increased hazard (HR, 1.78). More than 35% of the patients with AC survived past 7 years, whereas only 15% of the patients with SCC survived past this time. Controlling for grade confounded the histologic hazard relationship (HR, 1.68; CI, 0.99, 2.28) but was not significant. The elevated hazard for disease recurrence (HR, 4.16) significantly limits survival. Approximately 38% of the patients without recurrence survived past 7 years, whereas none of the patients with recurrence survived past this time. Although the curves indicate adjuvant treatment limits survival, the findings (HR, 1.82) were not significant. The sex-stratified curves in Figure 3 show that male patients have an elevated hazard (HR, 2.03) mainly for the short and midterm. Past 7 years, survival appears similar for both sexes (ie, about 26% of men and 26% of women survived). The DNA repair protein expression survival findings are shown in Supplemental Table 3 and were not significant. None of the DNA measures from the cosegregation analysis showed significance in HRs or statistical tests (not shown). Similarly, none of the DNA measures showed significance when restricting the analysis to patients with either the SCC or AC subtype (not shown) or by first removing the patients who underwent adjuvant treatment before performing the analysis (not shown).

Additional survival analysis was applied to evaluate various strata based on the stage I subgroups (graphs not shown), and the findings are provided in Supplemental Table 4. Stratification IA and IB produced little. However stratification by considering all patients in the lower age group with stage IA disease compared with the remaining population was associated with a significant hazard (HR, 2.44), indicating that older age in combination with stage IB disease confers poor survival relative to the remaining population with stage I disease. Similarly, stratification by considering patients with both AC and stage IA disease compared with the remaining population was significant (HR, 2.18). Stratification by considering patients with AC, stage IA disease, and lower age compared with the remaining population was significant (HR, 2.65). Thus, stage subgroup was related to survival when considering specific clinical factors of the population.

Discussion

The favorable outcome modeling showing that younger age and female sex were associated with a favorable survival is in agreement with previous reports.^{1,4,13} Our analysis also demonstrated that the histologic subtype in combination with sex and recurrence was associated with the greatest predictive value for a subset of patients. We also found that increasing age is related to SCC histologic subtype and increasing tumor grade is related to male sex. The survival analysis (Kaplan-Meier and Cox regression) clearly showed that

younger age, AC histologic subtype, no disease recurrence, and female sex confer longer survival. Differences within stage I subgroups were not directly related to survival. However, when considering certain clinical factors with stage IA, survival prospects are better. Younger patients with both the AC histologic subtype and stage IA disease have superior survival outcome. Our work showed that tumor grade was not a significant predictor of favorable outcome in contrast to other findings.¹ The increased hazard for patients with SCC in comparison with AC and for male patients and increasing age were significantly greater than those found in related work,¹ which may be due to either population or time frame differences. The survival findings with adjuvant therapy are consistent with a recent meta-analysis that documented an increased HR for adjuvant chemotherapy in patients with stage IA NSCLC,¹⁴ although our findings showed only a trend. The PARP and Ku86 measures showed little significant association with outcome, survival, or relationships with other variables, with the exception that Ku86 expression showed a significant association with male sex. DNA repair capacity has been shown to be a prognostic factor in patients with resected NSCLC. Patients with resected NSCLC whose tumors showed high *ERCC1* expression had a more favorable prognosis and failed to derive any benefit from adjuvant cisplatin-based chemotherapy.⁷ Other proteins involved in DNA damage repair mechanisms, such as Ku86 and PARP, have not been studied adequately in NSCLC.

None of the factors or combination of factors provided significant associations with recurrence or adjuvant chemotherapy. The value of recurrence in predicting favorable (or unfavorable) outcome or survivability may be limited because about 64% of the patients in the incident group and 92% of patients in the censored group did not have recurrence. In randomized studies for early-stage NSCLC, adjuvant chemotherapy improved survival for patients with stage II and stage IIIA disease,¹⁵ whereas patients with stage IA disease derived little benefit, in agreement with our findings.

The favorable outcome analysis supports patient level estimates and the dichotomization technique complements the time-to-event analysis. As demonstrated, a variable that provided a significant OR in the favorable outcome analysis also resulted in a significant hazard relationship when dichotomizing the same patient samples at the median value of the respective variable. A similar argument applies to a variable that gives an elevated Az (> 0.65) derived without model fitting as demonstrated previously.¹² This approach also serves as a simplifying technique for automated analyses.¹²

There are several limitations with our findings, the most significant of which is the retrospective data collection. This limitation resulted in incomplete case ascertainment such that the analytic data samples differed between various evaluations. In the modeling, we did not consider interaction terms to limit the presentation. We were able to construct LR models with increased predictive capability by limiting the work to 2 histologic subtypes, which limits the model's applicability. The recurrence status is not known past the censoring time for the relevant patients and represents an inexact or coarse variable.

The favorable outcome dichotomization analysis is a novel separation methodology.¹² This approach reduces uncertainty in the status for those patients who did not survive, but there are likely patients in the unfavorable group who survived longer than some patients in the

censored group. Assuming that the patients in the censored group did not die the day after losing study contact, their censored time is a conservative estimate (ie, left limit) of their overall survival time, indicating that the time separation (mean censored and incident times) between the 2 groups is greater than that specific to the separation of the censored group and incident group means. This indicates that associations found (ORs or Az) are more likely conservative estimates. Another approach¹¹ is to use a survival time cut point to dichotomize the patient population (ie, no possibility of overlapping survival time). This approach cannot accommodate censored patients on the left side of the cut point (ie, censored patients are discarded), which is not practical for limited data sets. The generality of our approach will require further evaluation with different data sets.

Understanding the factors that influence survival for patients with lung cancer remains an important research and clinical consideration. Evidence shows that the histologic distribution of lung cancer continues to change over time; specifically the rates of SCC cases have declined, whereas those for AC continue to rise.⁵ Lung cancer incidence rates for female patients appear to be rising.¹⁶ Moreover, the incidence rates vary regionally, with the highest incidence in the southern United States and the lowest incidence in the western United States.⁶ It follows that associations measured over one time frame or from a given region may not translate to another time frame or region, suggesting that they require continual assessments.

In conclusion, the survival characteristics of patients with stage I NSCLC were investigated. Our work showed that specific patient and tumor characteristics such as sex, age, and histologic subtype confer improved survival probability. These findings are of importance to determine optimal therapy and level of aggression required to manage stage I NSCLC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ms Candace Chisolm for her help with the tissue microarray construction and data digitization. We thank Ms Dianne Alexis for performing the immunohistochemical stains.

Supported by NIH P01 CA166999. GLS, FRK, TKO, and SSR are recipients of the Distinguished Cancer Scholar award from the Georgia Cancer Coalition.

References

1. Ou SH, Zell JA, Ziogas A, et al. Prognostic factors for survival of stage I nonsmall cell lung cancer patients: a population-based analysis of 19,702 stageI patients in the California cancer registry from 1989 to 2003. *Cancer*. 2007; 110:1532–41. [PubMed: 17702091]
2. Manser RL, Irving LB, Byrnes G, et al. Screening for lung cancer: a systematic review and meta-analysis of controlled trials. *Thorax*. 2003; 58:784–9. [PubMed: 12947138]
3. Aberle DR, Adams AM, Berg CD, et al. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365:395–409. [PubMed: 21714641]
4. Montesinos J, Bare M, Dalmau E, et al. The changing pattern of non-small cell lung cancer between the 90 and 2000 decades. *Open Respir Med J*. 2011; 5:24–30. [PubMed: 21754973]

5. Devesa SS, Bray F, Vizcaino AP, et al. International lung cancer trends by histologic type: male:female differences diminishing and adenocarcinoma rates rising. *Int J Cancer*. 2005; 117:294–9. [PubMed: 15900604]
6. Centers for Disease Control and Prevention (CDC). Racial/ethnic disparities and Geographic differences in lung cancer incidence—38 states and the district of Columbia, 1998–2006. *MMWR Morb Mortal Wkly Rep*. 2010; 59:1434–8. [PubMed: 21063273]
7. Olaussen KA, Dunant A, Fouret P, et al. DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *N Engl J Med*. 2006; 355:983–91. [PubMed: 16957145]
8. Patz EF Jr, Swensen SJ, Herndon JE 2nd. Estimate of lung cancer mortality from low-dose spiral computed tomography screening trials: implications for current mass screening recommendations. *J Clin Oncol*. 2004; 22:2202–6. [PubMed: 15169809]
9. Pisters KM, Le Chevalier T. Adjuvant chemotherapy in completely resected non-small-cell lung cancer. *J Clin Oncol*. 2005; 23:3270–8. [PubMed: 15886314]
10. Smith, A., Anand, SS. Patient survival estimation with multiple attributes: adaptation of Cox's regression to give an individual's point prediction. Presented at the 5th International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-2000): A workshop at the 14th European Conference on Artificial Intelligence (ECAI-2000; August 20–25 2000; Berlin, Germany.
11. Eschrich S, Yang I, Bloom G, et al. Molecular staging for survival prediction of colorectal cancer patients. *J Clin Oncol*. 2005; 23:3526–35. [PubMed: 15908663]
12. Behera M, Fowler EE, Owonikoko TK, et al. Statistical learning methods as a preprocessing step for survival analysis: evaluation of concept using lung cancer data. *Biomed Eng Online*. 2011; 10:97. [PubMed: 22067671]
13. Albain KS, Crowley JJ, LeBlanc M, et al. Survival determinants in extensive-stage non-small-cell lung cancer: the southwest Oncology Group experience. *J Clin Oncol*. 1991; 9:1618–26. [PubMed: 1651993]
14. Früh M, Rolland E, Pignon JP, et al. Pooled analysis of the effect of age on adjuvant cisplatin-based chemotherapy for completely resected non-small-cell lung cancer. *J Clin Oncol*. 2008; 26:3573–81. [PubMed: 18640938]
15. Pignon JP, Tribodet H, Scagliotti GV, et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE collaborative group. *J Clin Oncol*. 2008; 26:3552–9. [PubMed: 18506026]
16. Janssen-Heijnen ML, Coebergh JW. Trends in incidence and prognosis of the histological subtypes of lung cancer in North America, Australia, New Zealand and Europe. *Lung Cancer*. 2001; 31:123–37. [PubMed: 11165391]

Clinical Practice Points

- Approximately a third of patients with stage I NSCLC will experience recurrence of disease despite optimal surgical resection.
- It is important to identify patients at a higher risk for recurrence to develop novel treatment approaches to improve their outcome.
- We analyzed 161 patients with stage I NSCLC for prognostic factors associated with a favorable outcome.
- In addition to the baseline patient characteristics, the expression of Ku86 and PARP, 2 important DNA repair proteins, were studied as potential predictors of outcome.
- Male sex, squamous histologic sub type, and older age were associated with poorer survival outcomes, but the DNA repair proteins were not significant determinants.
- Our study confirms the favorable prognosis associated with female sex in NSCLC.
- Because new agents that target DNA repair pathways are under development, the findings of our study are relevant to researchers and clinicians.

Figure 1. Age Survival. This Shows the Survival Probability Curves for the Full (full group) Data Set. The Lower Age Group (upper blue curve) and Upper Age Group (lower brown curve) Were Formed by Using the Median Age as the Cut Point. The Lower Age Group Has a Survival Advantage (hazard ratio, 1.86)

Figure 2. Histologic Subtype Survival. This Shows the Survival Probability Curves for Patients With Either Adenocarcinoma (upper blue curve) or Squamous Cell Carcinoma (lower brown curve) Histologic Sub Type (subgroup 2). The Patients With Adenocarcinoma Clearly Have a Survival Advantage (hazard ratio, 1.7)

Figure 3.

Survival by Sex. This Shows the Survival Curves for Male Patients (brown) and Female Patients (blue) for the Full Group. Male Patients Have a Significantly Elevated Hazard (hazard ratio, 2.03) Compared With the Female Patients. The Survival Advantage for Female Patients Appears Limited to the Short and Midterm

Table 1

Patient Characteristics^a

Characteristic	I, N	I Mean/SD or %	C, N	C Mean/SD or %	Total, N	Total Mean/SD or %
Age						
	61	69.5/7.7	101	65.7/8.5	162	67.1/8.4
Grade						
	61	2.2/0.7	101	2.1/0.7	162	2.1/0.7
1	6	9.8%	19	18.8%	25	15.4%
2	35	57.4%	53	52.5%	88	54.3%
3	20	32.8%	29	28.7%	49	30.2%
Sex						
Male	39	63.9%	39	38.6%	78	48.1%
Female	22	36.1%	62	61.4%	84	51.8%
Histologic Subtype						
Adenocarcinoma	30	49.2%	63	62.4%	93	57.4%
Adenosquamous	2	3.3%	2	2.0%	4	2.5%
Large cell	2	3.3%	11	10.9%	13	8.0%
Squamous	26	42.6%	22	21.8%	48	29.6%
Unknown	1	1.6%	3	3.0%	4	2.5%
Smoking Status						
Nonsmoker	12	19.7%	19	18.8%	31	19.1%
Smoker	47	77.0%	74	73.3%	121	74.7%
Unknown	2	3.3%	8	7.9%	10	6.2%
Recurrence						
Yes	20	32.8%	6	5.9%	26	16.0%
No	39	63.9%	93	92.0%	132	81.5%
Unknown	2	3.3%	2	2.0%	4	2.5%
Tumor Position						
Lower lobe	18	29.5%	26	25.7%	44	27.1%
Middle lobe	5	8.2%	5	4.9%	10	6.1%
Upper lobe	36	59.0%	63	62.4%	99	61.1%

Characteristic	I, N	I Mean/SD or %	C, N	C Mean/SD or %	Total, N	Total Mean/SD or %
Upper/lower lobes	0	0.00%	2	2.0%	2	1.2%
Upper/middle lobes	0	0.00%	1	1.0%	1	0.6%
Chest wall	0	0.00%	1	1.0%	1	0.6%
Main stem bronchus	0	0.00%	1	1.0%	1	0.6%
Unknown	2	3.3%	2	2.0%	4	2.5%
Treatment						
Yes	10	16.4%	9	8.9%	19	11.7%
No	51	83.6%	92	91.0%	143	88.3%
Stage I						
A	37	60.6%	73	72.3%	110	67.9%
B	24	39.3%	28	27.7%	52	32.1%
K _p	61	96.9/6.6	101	97.1/3.9	162	97.0/5.1
K _s	61	289.4/24.8	101	288.3/21.5	162	288.7/22.8
P _I	61	2.5/0.6	101	2.5/0.6	162	2.5/0.6
P _p	61	83.5/17.5	101	84.9/18.2	162	84.4/17.9
P _s	61	223.4/76.2	101	221.5/77.4	162	222.3/76.7

^aThis table provides the patient characteristics for the incident group (I), censored group (C), and totals. The number of samples (N), mean values, standard deviation (SD), and percentages (%) are provided for each characteristic where applicable. The C and I groups correspond to the favorable and unfavorable outcome groups, respectively. The DNA expression measures for KU86 and PARP for the intensity, proportions and total score are A, K_I, K_p, K_s and P_I, P_p, and P_s, respectively.

Table 2

Full Group Associations^a

Model	Covariate	Unit/Reference	vs.	Covariate OR	Az
LR(A) = Censored Group	Age	8.4265	NA	0.61 (0.43, 0.87)	0.630
LR(G) = Censored Group	Sex	Female	Male	0.36 (0.18, 0.69)	0.627
LR(T) = Censored Group	Adjuvant treatment	No	Yes	0.50 (0.19, 1.31)	0.537
LR(Gr) = Censored Group	Grade	1.0000	NA	0.74 (0.45, 1.20)	0.549
LR(SI) = Censored Group	Stage I disease	A	B	0.59 (0.30, 1.16)	0.558
LR(Kp) = Censored Group	Kp	5.1212	NA	1.03 (0.75, 1.41)	0.479
LR(Ks) = Censored Group	Ks	22.7532	NA	0.95 (0.68, 1.32)	0.530
LR(Pi) = Censored Group	Pi	0.6120	NA	0.90 (0.65, 1.25)	0.527
LR(Pp) = Censored Group	Pp	17.8968	NA	1.08 (0.79, 1.47)	0.540
LR(Ps) = Censored Group	Ps	76.7112	NA	0.98 (0.71, 1.34)	0.498
LR(A,S) = Censored Group	Age	8.4265	NA	0.64 (0.44, 0.91)	0.683
	Sex	Female	Male	0.39 (0.20, 0.75)	

The DNA expression measures for KU86 and PARP for the intensity, proportions and total score are A, Ki, Kp, Ks and Pi, Pp, and Ps, respectively.

^aThis model included age (A), sex (S), adjuvant treatment (T) tumor grade (Gr), stage I subgroup (SI), Histologic sub type (H) restricted to adenocarcinoma (AC) and SCC, tumor location (loc), and DNA expression measures. The step-forward selection procedure was used to build a model to predict the censored group (favorable group). Only sex and the 2 histologic subtype variables were significant (univariate models). The odds ratios (ORs) with 95% confidence intervals and area under the receiver operating characteristic curve (Az) are provided for the various arrangements. We use the functional notation LR(x, y) = censored group to indicate the variables (ie, the respective covariates listed in the table) within the LR model used to predict the censored group. Entries that are not applicable are marked NA.

Table 3

Subgroup 1 Associations^a

Model	Covariate	Unit/Reference	vs.	Covariate OR	Az
LR(A) = Censored Group	Age	8.3671	NA	0.65 (0.45, 0.92)	0.615
LR(S) = Censored Group	Sex	Female	Male	0.32 (0.16, 0.64)	0.638
LR(T) = Censored Group	Adjuvant treatment	No	Yes	0.46 (0.17, 1.25)	0.542
LR(Gr) = Censored Group	Grade	1.0000	NA	0.52 (0.49, 1.36)	0.531
LR(S1) = Censored Group	Stage I	A	B	0.67 (0.33, 1.35)	0.544
LR(H) = Censored Group	Histologic sub type	Adenocarcinoma	Adenosquamous	0.50 (0.07, 3.73)	0.615
		Adenocarcinoma	Large cell	2.25 (0.46, 11.10)	
		Adenocarcinoma	Squamous	0.44 (0.21, 0.90)	
LR(loc) = Censored Group	Location	Lower lobe	Middle lobe	0.72 (0.18, 2.86)	0.537
		Lower lobe	Upper lobe	1.26 (0.60, 2.62)	
LR(Kp) = Censored Group	Kp	5.1312	NA	1.07 (0.77, 1.48)	0.488
LR(Ks) = Censored Group	Ks	23.0071	NA	0.98 (0.70, 1.37)	0.510
LR(Pt) = Censored Group	Pt	0.6026	NA	0.86 (0.61, 1.21)	0.535
LR(Pp) = Censored Group	Pp	17.6280	NA	1.01 (0.73, 1.41)	0.521
LR(Ps) = Censored Group	Ps	74.9729	NA	0.93 (0.66, 1.29)	0.514
LR(A, S) = Censored Group	Age	8.3671	NA	0.69 (0.48, 1.00)	0.680
	Sex	Female	Male	0.36 (0.18, 0.72)	

The DNA expression measures for KU86 and PARP for the intensity, proportions and total score are A, Kt, Kp, Ks and Pt, Pp, and Ps, respectively.

^aThis model included age (A), sex (S), adjuvant treatment (T), tumor grade (Gr), stage I subgroup (S1), Histologic sub type (H), restricted to adenocarcinoma (AC) and SCC Histologic sub type (H), tumor location (loc), and the DNA expression measures. The step-forward selection procedure was used to build a model to predict the censored group (the favorable outcome). The odds ratios (ORs) with 95% confidence intervals, and area under the receiver operating characteristic curve (Az) are provided for the various arrangements. The model with the best predictive capability is shown at the bottom using sex (S), Histologic sub type (H), and recurrence (Rec) and Histologic sub type restricted to AC and SCC. We use the functional notation LR(x, y) = censored group to indicate the variables (ie, the respective covariates listed in the table) within the LR model used to predict the censored group. Entries that are not applicable are marked NA.

Table 4

Subgroup 2 Associations^a

Model	Covariate	Unit/Reference	vs.	Covariate OR	Az
LR(A) = Censored Group	Age	7.9679	NA	0.71 (0.49, 1.02)	0.591
LR(S) = Censored Group	Sex	Female	Male	0.34 (0.17, 0.70)	0.630
LR(T) = Censored Group	Adjuvant treatment	No	Yes	0.41 (0.14, 1.22)	0.546
LR(S1) = Censored Group	Stage I	A	B	0.46 (0.21, 0.97)	0.582
LR(Gr) = Censored Group	Grade	1.0000	NA	0.73 (0.42, 1.26)	0.550
LR(H) = Censored Group	Histologic sub type	AC	SCC	0.44 (0.21, 0.91)	0.594
LR(loc) = Censored Group	Location	Lower lobe	Middle lobe	0.82 (0.20, 3.28)	0.547
		Lower lobe	Upper lobe	1.40 (0.65, 3.00)	
LR(K_P) = Censored Group	K _P	3.6013	NA	0.90 (0.63, 1.29)	0.502
LR(K_S) = Censored Group	K _S	19.6776	NA	0.82 (0.55, 1.23)	0.511
LR(P_I) = Censored Group	P _I	0.6058	NA	0.92 (0.65, 1.31)	0.515
LR(P_P) = Censored Group	P _P	18.0820	NA	1.04 (0.73, 1.46)	0.549
LR(P_S) = Censored Group	P _S	75.7872	NA	0.97 (0.69, 1.37)	0.493
LR(S) = Censored Group	Sex	Female	Male	0.34 (0.17, 0.70)	0.630
LR(S, H) = Censored Group	Sex	Female	Male	0.37 (0.18, 0.77)	0.668
	Histologic sub type	AC	SCC	0.49 (0.23, 1.04)	

The DNA expression measures for KU86 and PARP for the intensity, proportions and total score are A, K_I, K_P, K_S and P_I, P_P, and P_S, respectively.

^aThis model included age (A), sex (S), adjuvant treatment (T) tumor grade (Gr), stage I subgroup (S1), Histologic sub type (H) restricted to adenocarcinoma (AC) and SCC, tumor location (loc), and DNA expression measures. The step-forward selection procedure was used to build a model to predict the censored group (favorable group). Only sex and the 2 histologic subtype variables were significant (univariate models). The odds ratios (ORs) with 95% confidence intervals and area under the receiver operating characteristic curve (Az) are provided for the various arrangements. We use the functional notation LR(x, y) = censored group to indicate the variables (ie, the respective covariates listed in the table) within the LR model used to predict the censored group. Entries that are not applicable are marked NA.

Table 5

Best Model Associations^a

Model	Covariate	Unit/Reference	vs.	Covariate OR	Az
LR(A) = Censored Group	Age	7.9632	NA	0.72 (0.50, 1.05)	0.589
LR(S) = Censored Group	Sex	Female	Male	0.35 (0.16, 0.73)	0.630
LR(T) = Censored Group	Adjuvant treatment	No	Yes	0.36 (0.11, 1.15)	0.551
LR(S1) = Censored Group	Stage I	A	B	0.47 (0.21, 1.02)	0.580
LR(Gr) = Censored Group	Grade	1.0000	NA	0.71 (0.40, 1.25)	0.554
LR(H) = Censored Group	Histologic sub type	AC	SCC	0.46 (0.22, 0.98)	0.587
LR(loc) = Censored Group	Location	Lower lobe	Middle lobe	0.90 (0.22, 3.63)	0.544
		Lower lobe	Upper lobe	1.40 (0.63, 3.12)	
LR(K _P) = Censored Group	K _P	3.6995	NA	0.89 (0.62, 1.30)	0.499
LR(K _S) = Censored Group	K _S	19.6269	NA	0.85 (0.56, 1.27)	0.502
LR(P _I) = Censored Group	P _I	0.6208	NA	0.91 (0.63, 1.31)	0.513
LR(P _P) = Censored Group	P _P	18.4569	NA	1.04 (0.73, 1.48)	0.541
LR(P _S) = Censored Group	P _S	77.3386	NA	0.97 (0.68, 1.39)	0.498
LR(Sm) = Censored Group	Smoking	No	NA	1.03 (0.44, 2.42)	0.503
LR(Rec) = Censored Group	Recurrence	No	NA	0.06 (0.01, 0.25)	0.659
LR(S, H, Rec) = Censored Group	Sex	Female	Male	0.32 (0.14, 0.76)	0.788
	Histologic sub type	AC	SCC	0.41 (0.17, 0.98)	
	Recurrence	No	Yes	0.04 (0.01, 0.20)	
LR(A, S, H, Rec) = Censored Group	Age	7.9632	NA	0.73 (0.47, 1.15)	0.796
	Sex	Female	Male	0.35 (0.15, 0.84)	
	Histologic sub type	AC	SCC	0.45 (0.19, 1.08)	
	Recurrence	No	Yes	0.04 (0.01, 0.19)	

The DNA expression measures for KU86 and PARP for the intensity, proportions and total score are A, K_I, K_P, K_S and P_I, P_P, and P_S, respectively.

^aThis model included age (A), sex (S), adjuvant treatment (T), tumor grade (Gr), stage I subgroup (S1), Histologic sub type (H), restricted to adenocarcinoma (AC) and SCC Histologic sub type (H), tumor location (loc), and the DNA expression measures. The step-forward selection procedure was used to build a model to predict the censored group (the favorable outcome). The odds ratios (ORs) with 95% confidence intervals, and area under the receiver operating characteristic curve (Az) are provided for the various arrangements. The model with the best predictive capability is shown at the bottom using sex

Author Manuscript   Author Manuscript   Author Manuscript   Author Manuscript

(S), Histologic sub type (H), and recurrence (Rec) and Histologic sub type restricted to AC and SCC. We use the functional notation $LR(x, y)$ = censored group to indicate the variables (ie, the respective covariates listed in the table) within the LR model used to predict the censored group. Entries that are not applicable are marked NA.

Table 6

Survival Analysis Associations

Model/Group	n (n _i , n _c)	Wilcoxon χ^2 (P Value)	Log-Rank χ^2 (P Value)	HR (95% CI)	3 Yr % Survival	5 Yr % Survival	7 Yr % Survival
Survival Age	162 (61, 101)	5.01 (.0252)	5.75 (.0165)	1.86 (1.11, 3.12)			
Lower Age Group (Ref)	82 (24, 58)				79.11%	64.39%	32.20%
Upper Age Group	80 (37, 43)				63.28%	46.71%	29.12%
Survival Histologic Statistics	141 (56, 85)	5.08 (.0242)	4.68 (.0304)	1.78 (1.05, 3.01)			
AC (Ref)	93 (30, 63)				77.33%	57.64%	35.47%
SCC	48 (26, 22)				57.35%	45.53%	15.18%
Survival Recurrence	158 (59, 99)	18.28 (.0001)	28.79 (.0001)	4.16 (2.37, 7.31)			
No Recurrence (Ref)	132 (39, 93)				79.55%	64.97%	33.87%
Recurrence	26 (20, 6)				33.65%	11.22%	0.00%
Survival Treatment	162 (61, 101)	1.62 (.2030)	3.04 (.0811)	1.82 (0.92, 3.61)			
No Treatment (Ref)	143 (51, 92)				72.39%	59.82%	29.56%
Treatment	19 (10, 8)				63.16%	28.07%	28.07%
Survival Sex	162 (61, 101)	9.39 (.0022)	7.27 (.0070)	2.03 (1.20, 3.43)			
Female (Ref)	84 (22, 62)				81.72%	67.55%	25.80%
Male	78 (39, 39)				60.23%	44.75%	26.85%

This table provides the hazard ratios (HRs) with 95% confidence intervals, the Wilcoxon, χ^2 , and log-rank test P-values, and the percentage of patients surviving past 3, 5, and 7 years for the various groups. The number of patients in each stratification belonging to the censored group (n_c), incident group (n_i), and totals (N) for each experiment are also provided. We show the survival statistics for age, histologic subtype restricted to adenocarcinoma (AC) and SCC disease recurrence (Recurrence), adjuvant treatment (Treatment), and sex. The reference (Ref) groups are designated in the table.