Published in final edited form as:

Lung Cancer. 2015 November ; 90(2): 281–287. doi:10.1016/j.lungcan.2015.08.007.

Validation of Survival Prognostic Models for Non-Small-Cell Lung Cancer in Stage- and Age-specific Groups

Xiaofei Wang¹, Lin Gu¹, Ying Zhang¹, Daniel J. Sargent², William Richards³, Apar Kishor Ganti⁴, Jeffery Crawford⁵, Harvey Jay Cohen⁵, Thomas Stinchcombe⁶, Everett Vokes⁷, and Herbert Pang^{1,8}

¹Department of Biostatistics & Bioinformatics and Alliance Statistics and Data Center, Duke University, Durham, NC; supported by U10CA180882, CA033601 and NIA5R21AG042894

²Alliance Statistics and Data Center, Mayo Clinic, Rochester, MN; supported by U10CA180882 and CA033601

³Brigham and Women's Hospital, Boston, MA

⁴Department of Internal Medicine, VA Nebraska Western Iowa Health Care System and University of Nebraska Medical Center, Lincoln, NE; supported by CA180821 and VA Career Development Award

⁵Department of Medicine, Duke University Medical Center, Durham, NC; supported by CA47577

⁶Department of Medicine, University of North Carolina, Chapel Hill, NC; supported by CA47559 and NIA 5R21AG042894

⁷Department of Medicine, University of Chicago, Chicago, IL; supported by CA41287 and NIA 5R21AG042894

⁸School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; supported by NIA 5R21AG042894

Abstract

Purpose—Prognostic models have been proposed to predict survival for non-small-cell lung cancer (NSCLC). It is important to evaluate whether these models perform better than performance status (PS) alone in stage- and age-specific subgroups.

Patients and Methods—The validation cohort included 2060 stage I and 1611 stage IV NSCLC patients from 23 CALGB studies. For stage I, Blanchon (B), Chansky (C) and Gail (G)

Correspondence: Xiaofei Wang, Ph.D., Duke University, 2424 Erwin Road Suite 1102, Durham, NC 27705, 919-681-5406, xiaofei.wang@duke.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Presented in part in abstract form at the 104th Annual Meeting of the American Association for Cancer Research, Washington, DC, April 6-10, 2013.

CONFLICT OF INTEREST DISCLOSURES

There are no conflicts of interest.

models were evaluated along with the PS only model. For stage IV, Blanchon (B) and Mandrekar (M) models were compared with the PS only model. The c-index was used to assess the concordance between survival and risk scores. The c-index difference (c-difference) and the integrated discrimination improvement (IDI) were used to determine the improvement of these models over the PS only model.

Results—For stage I, B and PS have better survival separation. The c-index for B, PS, C and G are 0.61, 0.58, 0.57 and 0.52, respectively, and B performs significantly better than PS with c-difference=0.034. For stage IV, B, M and PS have c-index 0.61, 0.64 and 0.60, respectively; B and M perform significantly better than PS with c-difference=0.015 and 0.033, respectively.

Conclusion—Although some prognostic models have better concordance with survival than the PS only model, the absolute improvement is small. More accurate prognostic models should be developed; the inclusion of tumor genetic variants may improve prognostic models.

Keywords

Independent validation; Non-small cell lung cancer (NSCLC); Performance status; Prognostic models

Introduction

Lung cancer is the leading cause of cancer death among both men and women in the United States [1]. Over 85% of lung cancer cases are non-small cell lung cancer (NSCLC) [2]. Performance status (PS) is a simple functional assessment based on daily physical activities. There are two mutually convertible scoring systems for PS: the Karnofsky score [3] and the ECOG score, also called Zubrod/WHO score [4]. While PS is criticized for its subjectivity [5] and lack of prediction for chemotherapy toxicity [6], PS has become a popular prognostic tool in practice [7] and is one of the most commonly used eligibility criteria and stratification factors in randomized trials [8].

Even with the success of PS, the lung cancer community has been keen on developing new, potentially more objective and accurate prognostic models. Four existing prognostic models for NSCLC patients are summarized in Table 1. It is not surprising that three of the four models include PS as one of the predictors. Blanchon et al. [9] developed a prognostic model (B model) using the data from 2979 NSCLC patients. The prognostic model was based on multivariate Cox regression modeling with baseline prognostic factors: age, histology, PS, sex and stage (I-IV). Chansky et al. [10] (C model) assembled data from 9137 surgically resected NSCLC patients from North America, and used both Cox regression and recursive partitioning and amalgamation analyses to identify risk factors, including age, sex and stage (I-III A), and provide classification of risk groups. A prognostic model established by Gail et al. [11] (G model) using 392 early stage NSCLC patients. The prognostic model was based on a Weibull survival model with histology, PS, TNM staging and post-operative infections (empyema, pneumonia or wound infections) as risk predictors. Mandrekar et al. [12] developed a prognostic model (M model) using 782 advanced NSCLC patients. This model incorporated patient characteristics, such as age, PS, sex, stage and body mass index

(BMI), and pre-treatment laboratory values, such as hemoglobin (HGB) and white blood cell (WBC) count.

However, these models are not commonly used in clinical practice or clinical trial design. Compared to the PS only model, these models require additional prognostic factors and some factors (e.g. lab values) may not be readily available. Most of these models already include PS as a predictor in the statistical models, but the discriminative accuracies of these models relative to the PS only model have not yet been fully validated using data from independent studies.

Since 1988, the Cancer Leukemia Group B (CALGB) (now part of the Alliance for Clinical Trials in Oncology) has conducted phase II and III clinical trials with NSCLC patients. The goal of the present research was to evaluate the accuracy of the four externally developed prognostic models noted above, for predicting overall survival of stage I and stage IV NSCLC patients. The added value of these prognostic models in discriminating overall survival over the PS only model was a focus. Also, as these prognostic models were predominantly developed on young patients (defined as age <70 years), we were also interested in evaluating the performance of these prognostic models among elderly patients (defined as age ≥ 70 years).

Patients and Methods

Selection Criteria

This study utilized data from the CALGB clinical trials to validate four published prognostic models in stage I and stage IV NSCLC. We identified all NSCLC studies targeting either stage I or IV patients conducted between 1988 and 2009 by the CALGB. Stage II and III patients were excluded because of the limited number of patients. Ongoing trials and recently closed trials with pending publications, trials with missing staging, diagnosis and histology information were excluded. The 23 CALGB studies included in the validation analysis are listed in the supplementary material Table S1. Of note none of the trials included selected patients based on epidermal growth factor receptor (*EGFR*) mutation or anaplastic lymphoma kinase (*ALK*) rearrangement status. The validation cohort consists of 2060 stage I (IA-IB) and 1611 stage IV NSCLC patients, among which 2453 patients were <70 years old and 1218 patients were ≥ 70 years old.

Prognostic Models

Calculations of the risk scores based on the four prognostic models are briefly summarized in Table 2. Details on variable definitions and risk score calculations can be found in the corresponding publications [9-12]. Risk groups generating potential separation of survival curves can be defined on mutually exclusive intervals of risk scores. Six risk groups for the B model are formed on the score ranges specified in Table 3 of Blanchon et al [9]. Five risk groups for the C model are formed based on the survival tree in Figure 2 of Chansky et al [10]. Three risk groups for the G model are formed based on the rule in Table 9 of Gail et al [11]. Risk group was not discussed in Mandrekar et al [12]. For illustration, we created five risk groups for the M model by grouping patients at the quintiles of the risk score.

Statistical considerations for validation

TNM information captured at patient registration was used to determine staging according to AJCC version 6; Version 7 was not used because available information is not sufficient to convert all patients from version 6 stage to version 7 stage. If both pathological and clinical staging were available, pathological staging was used. Complete cases analysis was adopted as the primary method to handle missing variables; specifically patients with missing predictors were excluded from the validation cohort. Sensitivity analyses with multiple imputation (MI) to handle missing variables were also performed. The MI analysis was based on 5 imputed datasets and the estimates were combined using Rubin's method [36].

For stage I the B, C and G models were evaluated along with the PS only model. For stage IV, the PS, B and M models were evaluated. The assessments were conducted for young (defined as age < 70 years), elderly (defined as age ≥ 70 years), and all patients combined.

The outcome of interest was overall survival, defined as the time from registration to the time of death. Overall survival was censored at the date of last follow-up, when death was not observed. The survival curves between risk groups were compared using the Kaplan-Meier method and the log-rank test. Hazard ratios for the different risk scores were calculated using a Cox proportional hazards model. Assuming 40% and 90% death rate for stage I and stage IV patients, respectively, in the validation cohort, we have at least 95% power to a hazard ratio 1.20 for a risk score with a standard deviation of 1.0 at a 2-sided significance level of 0.10 for both patient groups.

To ease comparison, all risk scores are standardized to scores between 0 and 10. We estimated c-index (c) for a global assessment of the concordance between risk score and survival [37]. The difference in c-index (c-difference) is a measure formed by subtracting the c-index of the PS model from the c-index of a risk score. The c-difference and its bootstrap confidence interval based on 1000 bootstrap samples were used to evaluate the improvement of concordance of a risk score over the PS only model. To facilitate the comparison of c-index estimates among different patient subgroups, a modified c-index, free of influence of censoring distribution was also estimated [38]. The c-difference is a rank based statistics and it has been show insensitive to detect small differences in the discrimination accuracy between predictive models. Integrated discrimination improvement (IDI) is a difference between improvement in average sensitivity and improvement in average (1-specificity) and has been proposed to address the insensitive issue c-index [39, 40]. By presenting both statistics, we attempt to evaluate the added value of a risk score over the standard model. All statistical analyses were performed using SAS 9.3 (Cary, NC) and R 3.0.1 (boot, rms, survC1 and survIDINRI packages).

Results

As seen in Table 3, for the 2060 stage I patients, the median age was 67.9 years (range 24 to 95), 53.5% were male, 86.1% white, 64.7% PS=0, and 11.2% with weight loss ≥ 5% in the past 3 or 6 months prior to enrollment. The most common histology was adenocarcinoma (50.4%) followed by squamous histology (30%). A total 914 deaths were observed among the 2060 patients with a median follow-up time of 4.9 years.

For the 1611 stage IV patients, the median age was 62.6 years (range 25 to 88), 63.4% were male, 83.1% white, 35.1% PS=0, and 34.2% had weight loss $\geq 5\%$ in the past 3 or 6 months prior to enrollment. The most common histology was adenocarcinoma (55.8%) followed by squamous (18.5). Of these stage IV patients, a total of 1513 deaths in the 1611 patients were observed with a median follow-up time of 3.3 years.

Stage I Prognostic Models

For all stage I patients, the median survival was 6.55 years (95%CI: 6.09-6.95) and the 5-year OS rate was 0.58 (95%CI: 0.56-0.61). The risk groups derived from the four prognostic models all yield statistically significant separation among the risk groups (log rank p-value <0.001). The risk groups of B, C and PS models correctly ordered the survival functions, in which higher risk groups yield worse survival (Figure 1), but the G model did not (Table 4). The risk score of the B model performed the best among all models in predicting survival of stage I patients. The hazard ratios for the B, C, G and PS models were 1.441, 1.126, 1.003 and 1.238, respectively (Table 4). The c-indices for the B, C, G and PS models were 0.611, 0.573, 0.519, and 0.580, respectively (Table 5). When the models were compared to the PS only model, the B model had a significant improvement in discriminating survival, as indicated by c-difference of 0.034 with a 95%CI (0.023-0.045) excluding zero, but not the integrated discrimination index (IDI) 0.029 (95%CI: -0.044-0.061) where the confidence interval included zero. Because 19% stage I patients in the validation cohort had missing risk scores for the C model, the multiple imputation (MI) analysis was used to evaluate the impact of the missing data on model evaluation. All performance measures from the MI analysis for the C model are very similar to those from the complete case analysis, for example, the hazard ratio and c-index are 1.13 and 0.58 for the MI analysis.

There is a statistically significant difference in survival between stage I young and old patients (log rank p-value <0.0001) with median OS of 7.88 years (95%CI: 7.28-8.94) for young and 4.92 years (95%CI: 4.49-5.63) for old patients. The results of age group-specific validation are provided in the supplementary figures and tables (Figures S1-S2, Tables S2-S4). The performances of the four prognostic models in young stage I patients, with c-indices for B, C, G and the PS models of 0.587, 0.548, 0.536 and 0.564, respectively, resemble closely those seen in the entire cohort. Similar findings hold true for older stage I patients in that the B model was still the best model and significantly better than the PS model. As indicated clearly by c-difference and IDI, C and G models were significantly worse than the PS only model. The incorrect survival ordering of the risk groups of C and G models suggests they are inappropriate to use in patient prognostication and stratification for stage I NSCLC. Overall, the B model has the best performance in all stage I patients, in young stage I patients, as well as in old stage I patients, but compared to the PS only model, the absolute improvement in concordance between the B model risk score and survival was limited, ranging from 2.44%-3.48%.

Stage IV Prognostic Models

For all stage IV patients, the median survival was 8.25 months (95%CI: 7.79-8.87) and the 6-month OS rate was 0.62 (95%CI: 0.60-0.65). The B, M and the PS models are applicable to stage IV patients. Because the predictors such as weight loss, WBC and HGB were

missing in approximately 30% of stage IV patients, the actual number of patients used in validating the M model was about 67% of the B model and the PS only model. Additional analysis was conducted using the subset of 1081 patients who have risk scores for both the B model and the M model. Because the findings are very similar, the results for patients with at least one risk score are reported here. The risk groups derived from the prognostic models yield statistically significant separation in survival (log rank p-value <0.001). The risk groups of all models produce a correct ordering of survival functions (Figure 1, Table 4). The c-indices for B, M and PS models were 0.611, 0.641 and 0.596, respectively (Table 5). Because 33% stage IV patients in the validation cohort had missing variables for the M model, the multiple imputations (MI) analysis was used to evaluate the impact of the missing data on model evaluation. All performance measures from the MI analysis are very similar to those from the complete case analysis, for example, the c-index from the MI analysis is 0.62 versus 0.61 from the complete case analysis. Both the B and M models performed better than the PS only model with c-difference=0.015 (95%CI: 0.008-0.023) and 0.033 (95%CI: 0.020-0.046), respectively, with similar improvement in IDI indices. There were no significant differences in OS between stage IV young and old patients (log rank p-value=0.347) with median OS 8.34 months (95%CI: 7.82-9.10) for young and 7.85 months (95%CI: 6.67-9.10) for older patients. As seen in the supplementary Figures S1-S2 and Tables S2-S4, similar performance patterns of these prognostic models hold in younger and older stage IV patients. Overall, the B and M models have significantly better discrimination than the PS only model in all stage IV patients, and in the younger and older subsets, but the absolute improvement in concordance between risks scores and survival is limited with a range of 2-5%.

Discussion

The objective of this analysis was to evaluate the accuracy of four externally developed prognostic models for predicting overall survival of stage I and stage IV NSCLC patients. We were particularly interested in the performance of these prognostic models in elderly lung cancer patients relative to younger patients, as elderly patients have been less investigated in clinical trials. Overall, the B model has the best prognostic capability for stage I and IV as well as younger and older patients. It uses easily measured covariates for computing risk scores and divides patients into 6 risk groups based on risk scores. For stage IV patients, the M model is particularly strong with the incorporation of additional variables, such as weight loss, WBC and HGB. One of the reasons why the C model may have performed poorly in stage I patients is that PS is not included in the model's definition of risk groups. The C model performs significantly worse than the PS only model in the elderly patients, and classifies the entire set of elderly patients as one risk group. The G model included similar predictors to the B model, including histology, stage and PS, with wound infection as a unique predictor. Unfortunately, in our validation cohort, infection data was not collected in about 80% patients of the stage I NSCLC trials; of 421 patients who had infection data recorded, there was only one patient reporting infection. For this reason, we assumed all stage I patients had no infection, or equivalently the regression coefficient of infection is zero, in the analysis. The relatively poor performance of the G model could be attributed to this and a few other reasons. For example, this model was developed on a

relatively small cohort (392 patients); it was not validated with internal or external datasets; and a parametric regression model was used to estimate risk score. It is interesting to note that most prognostic models using additional covariates fail to perform better for the PS only model for stage I, stage IV, old and young patients. The only exceptions are the B model in stage I and IV and the M model for stage IV, but the absolute improvement in concordance of B and M models over the PS model was still small, ranging from 2-5%, and IDI ranging from 0.001-0.029. Because of its simplicity and the small added value for the more complicated models for patients with stage I and stage IV disease, the PS only model may continue to be the most popular score for stratification in clinical trials. However, with the concordance of the PS model with survival being 0.58 for stage I and 0.60 for stage IV patients, PS alone remains a suboptimal prognostic model.

Prognostication of survival and adverse events is particularly important in elderly lung cancer patients, as these patients may be particularly vulnerable and have been less well studied than their younger counterparts. To date, prognostic models have been primarily developed using a younger patient population. As the results show, some of the existing models (e.g. B and G) are worse than the PS only model when used to predict survival in elderly patients. Comorbidity and other geriatric risk factors have previously shown a strong association with survival and chemotherapy-related toxicity in elderly lung cancer patients [6, 41]. None of these prognostic factors were used in the existing models, likely because these prognostic factors were not captured in the database used to develop these models. This may be why these models failed to provide a significant discriminative advantage over PS alone in elderly patients.

This study has limitations. Not every published prognostic model for NSCLC was compared. To avoid the issue of incomparability, prognostic model developed based on patients exclusively treated by one drug such as bevacizumab [42] was not chosen for comparison. While all relevant CALGB trials from 1988-2009 have been identified and included, the cohort of patients available for studying the elderly is still small. Most patients had good PS as required for trial eligibility which led to a small validation cohort for patients with poor PS. In the validation, like most of the prognostic models, we included patients who were treated by different surgical procedures, chemotherapy and radiation. Although it is important for prognostic models to predict outcome independent of the exact nature of the treatment paradigm, the heterogeneity introduced by treatment variation may have underestimated the absolute prediction accuracy of these models.

Given the limited success of the existing prognostic models, including PS alone model, there is a clear need for developing prognostic models that improve considerably upon PS alone, for patient stratification in stage- and age-specific clinical trials. Further research, including the application of geriatric assessment and frailty indicators, is needed to address these limitations, especially for the older and more vulnerable patients. Regardless, our study suggests that the predictive ability of existing more complex models is not meaningfully better than that of PS alone, particularly for older patients.

Lung cancer patients are at considerable risk for other causes of deaths due to their advanced age at diagnosis and smoking-related chronic diseases. The discriminative accuracy of a

prognostic model can be improved by taking the issue of competing risk into account. With the success of the identification of genetic alterations, such as *EGFR* mutations and *ALK* rearrangements, and development of targeted therapies for NSCLC, it has become clear that lung cancer is a heterogeneous disease. The prognostication and its treatment should incorporate the genetic profile of patients as well [43].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers to the Alliance for Clinical Trials in Oncology (U10CA180821 and U10CA180882) and the legacy CALGB (CA31946, CA33601, CA180821, CA47577, CA47559, and CA41287). Additional funding was provided by NIA 5R21AG042894 (X. W., L. G., Y. Z., T. S., E. V., and H. P.) and a VA Career Development Award (A.K.G.).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. CDC website on lung cancer. <http://www.cdc.gov/cancer/lung/>
2. Molina JR, Yang P, Cassivi SD, et al. Non-small cell lung cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clin Proc.* 2008; 83(5):584–594. [PubMed: 18452692]
3. Karnofsky, DA.; Burchenal, JH. The Clinical Evaluation of Chemotherapeutic Agents in Cancer.. In: MacLeod, CM., editor. *Evaluation of Chemotherapeutic Agents*. Columbia Univ Press; 1949. p. 196
4. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am. J. Clin. Oncol.* 1982; 5(6):649–55. [PubMed: 7165009]
5. Schag CC, Heinrich RL, Ganz PA. Karnofsky performance status revisited: reliability, validity, and guidelines. *J Clin Oncol.* 1984; 2(3):187–93. [PubMed: 6699671]
6. Hurria A, Togawa K, Mohile SG, et al. Predicting chemotherapy toxicity in older adults with cancer: a prospective multicenter study. *J Clin Oncol.* 2011; 29(25):3457–65. [PubMed: 21810685]
7. Ettinger DS, Akerley W, Borghaei H, et al. Non-small cell lung cancer, version 2.2013. *J Natl Compr Canc Netw.* 2013; 11(6):645–53. [PubMed: 23744864]
8. Sorbye H, Köhne CH, Sargent DJ, Glimelius B. Patient characteristics and stratification in medical treatment studies for metastatic colorectal cancer: a proposal for standardization of patient characteristic reporting and stratification. *Ann Oncol.* 2007; 18(10):1666–72. [PubMed: 17660495]
9. Blanchon F, Grivauux M, Asselain B, et al. 4-year mortality in patients with non-small-cell lung cancer: development and validation of a prognostic index. *Lancet Oncology.* 2006; 7:829–836. [PubMed: 17012045]
10. Chansky K, Sculier JP, Crowley JJ, et al. International Staging Committee and Participating Institutions. The International Association for the Study of Lung Cancer Staging Project: prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. *J Thorac Oncol.* 2009; 4:792–801. [PubMed: 19458556]
11. Gail MH, Eagan RT, Feld R, et al. the Lung Cancer Study Group. Prognostic factors in patients with resected stage I non-small cell lung cancer. *Cancer.* 1984; 54:1802, 1802–13. [PubMed: 6478416]
12. Mandrekar SJ, Schild SE, Hillman SL, et al. A prognostic model for advanced stage non-small cell lung cancer Pooled analysis of North Central Cancer Treatment Group trials. *Cancer.* 2006; 107:781–792. [PubMed: 16847887]

13. Vokes EE, Lyss AP, Herndon JE 2nd, et al. Intravenous 6-thioguanine or cisplatin, fluorouracil and leucovorin for advanced non-small cell lung cancer: a randomized phase II study of the cancer and leukemia group B. *Ann Oncol.* 1992; 3:727–32. [PubMed: 1333267]
14. Graziano SL, Herndon JE 2nd, Richards F 2nd, et al. A phase I trial of ifosfamide, mesna, and cisplatin in advanced non-small cell lung cancer. A cancer and leukemia group B study. *Cancer.* 1993; 72:62–8. [PubMed: 8389667]
15. Muscato JJ, Cirrincione C, Clamon G, et al. Etoposide (VP-16) and cisplatin at maximum tolerated dose in non-small cell lung carcinoma: A cancer and leukemia group B study. *Lung Cancer.* 1995; 13:285–294. [PubMed: 8719068]
16. Kosty MP, Fleishman SB, Herndon JE 2nd, et al. Cisplatin, vinblastine, and hydrazine sulfate in advanced, non-small-cell lung cancer: a randomized placebo-controlled, double-blind phase III study of the Cancer and Leukemia Group B. *J Clin Oncol.* 1994; 12:1113–20. [PubMed: 8201372]
17. Graziano SL, Valone FH, Herndon JE 2nd, et al. A randomized phase II study of ifosfamide/mesna/cisplatin plus G-CSF or etoposide/cisplatin plus G-CSF in advanced non-small cell lung cancer: a Cancer and Leukemia Group B study. *Lung Cancer.* 1996; 14(2-3):315–29. [PubMed: 8794413]
18. Loewen GM, Watson D, Kohman L, et al. Preoperative exercise Vo2 measurement for lung resection candidates: results of Cancer and Leukemia Group B Protocol 9238. *J Thorac Oncol.* 2007; 2:619–25. [PubMed: 17607117]
19. Shennib H, Bogart J, Herndon JE, et al. Video-assisted wedge resection and local radiotherapy for peripheral lung cancer in high-risk patients: the Cancer and Leukemia Group B (CALGB) 9335, a phase II, multi-institutional cooperative group study. *J Thorac Cardiovasc Surg.* 2005; 129:813–8. [PubMed: 15821648]
20. Perry MC, Ihde DC, Herndon JE 2nd, et al. Paclitaxel/ifosfamide or navelbine/ifosfamide chemotherapy for advanced non-small cell lung cancer: CALGB 9532. *Lung Cancer.* 2000; 28:63–8. [PubMed: 10704711]
21. Strauss GM, Herndon JE 2nd, Maddaus MA, et al. Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups. *J Clin Oncol.* 2008; 26:5043–51. [PubMed: 18809614]
22. Maddaus MA, Wang X, Vollmer RT, et al. CALGB 9761: A prospective analysis of IHC and PCR based detection of occult metastatic disease in stage I NSCLC. *J Clin Oncol.* Jun.2006 24(18_suppl 7030) Meeting Abstracts.
23. Akerley W, Herndon JE Jr. Lyss AP, et al. Induction paclitaxel/carboplatin followed by concurrent chemoradiation therapy for unresectable stage III non-small-cell lung cancer: a limited-access study--CALGB 9534. *Clin Lung Cancer.* 2005; 7:47–53. [PubMed: 16098244]
24. Lilenbaum RC, Herndon JE 2nd, List MA, et al. Single-agent versus combination chemotherapy in advanced non-small-cell lung cancer: the cancer and leukemia group B (study 9730). *J Clin Oncol.* 2005; 23:190–6. [PubMed: 15625373]
25. Swanson SJ, Herndon JE 2nd, D'Amico TA, et al. Video-assisted thoracic surgery lobectomy: report of CALGB 39802--a prospective, multi-institution feasibility study. *J Clin Oncol.* 2007; 25:4993–7. [PubMed: 17971599]
26. Rocha Lima CM. Randomized phase II trial of gemcitabine plus irinotecan or docetaxel in stage IIIB or stage IV NSCLC. *Annals of Oncology.* 2004; 15:410–418. [PubMed: 14998842]
27. Bogart JA, Hodgson L, Seagren SL, et al. Phase I study of accelerated conformal radiotherapy for stage I non-small-cell lung cancer in patients with pulmonary dysfunction: CALGB 39904. *J Clin Oncol.* 2010; 28:202–6. [PubMed: 19933904]
28. Edelman MJ, Watson D, Wang X, et al. Eicosanoid modulation in advanced lung cancer: cyclooxygenase-2 expression is a positive predictive factor for celecoxib + chemotherapy--Cancer and Leukemia Group B Trial 30203. *J Clin Oncol.* 2008; 26:848–55. [PubMed: 18281656]
29. Sugarbaker DJ, Richards WG, Wang XF, et al. THE CALGB LUNG CANCER TISSUE BANK. CALGB 140202 protocol. 2004
30. Miller AA, Wang XF, Gu L, et al. Phase II Randomized Study of Dose-Dense Docetaxel and Cisplatin Every 2 Weeks With Pegfilgrastim and Darbepoetin Alfa With and Without the

- Chemoprotector BNP7787 in Patients With Advanced Non-small Cell Lung Cancer (CALGB 30303). *Journal of Thoracic Oncology*. 2008; 3:1159–1165. [PubMed: 18827613]
31. Liptay MJ, D'Amico TA, Nwogu C, et al. Intraoperative sentinel node mapping with technitium-99 in lung cancer: results of CALGB 140203 multicenter phase II trial. *J Thorac Oncol*. 2009; 4:198–202. [PubMed: 19179896]
32. Lilenbaum R, Wang X, Gu L, et al. Randomized phase II trial of docetaxel plus cetuximab or docetaxel plus bortezomib in patients with advanced non-small-cell lung cancer and a performance status of 2: CALGB 30402. *J Clin Oncol*. 2009; 27:4487–91. [PubMed: 19704058]
33. Janne PA, Wang X, Socinski MA, et al. Randomized phase II trial of erlotinib alone or with carboplatin and paclitaxel in patients who were never or light former smokers with advanced lung adenocarcinoma: CALGB 30406 trial. *J Clin Oncol*. 2012; 30:2063–9. [PubMed: 22547605]
34. Heist RS, Wang X, Hodgson L, et al. A randomized phase II study to assess the efficacy of pemetrexed or sunitinib or pemetrexed plus sunitinib in the second-line treatment of advanced non-small-cell lung cancer. *J Thorac Oncol*. 2014; 9(2):214–21. [PubMed: 24419419]
35. Harpole DH, Wang XF, Kratzke R, et al. A Randomized Phase III Trial to Evaluate the Potential Utility of a Genomic Prognostic Model to Identify Stage I Non-Small Cell Lung Cancer (NSCLC) Patients as Candidates for Adjuvant Chemotherapy. CALGB 30506 Protocol. 2009
36. Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons; New York: 1987.
37. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat in Med*. 1996; 15:361–387. [PubMed: 8668867]
38. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*. 2011; 30:1105–16. [PubMed: 21484848]
39. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine*. 2011; 30(1):11–21. [PubMed: 21204120]
40. Uno H, Tian L, Cai T, Kohane IS, Wei LJ. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Statistics in Medicine*. 2013; 32(14):2430–42. [PubMed: 23037800]
41. Chen RC, Royce TJ, Extermann M, Reeve BB. Impact of age and comorbidity on treatment and outcomes in elderly cancer patients. *Semin Radiat Oncol*. 2012; 22(4):265–71. [PubMed: 22985808]
42. Hoang T, Dahlberg SE, Sandler AB, et al. Prognostic Models to Predict Survival in Non-Small Cell Lung Cancer Patients Treated with First-Line Paclitaxel and Carboplatin with or without Bevacizumab. *J Thorac Oncol*. 2012; 7(9):1361–1368. [PubMed: 22843087]
43. Pang H, Wang X. Statistical aspect of translational and correlative studies in clinical trials. *Chin Clin Oncol*. 2015 (Invited). In press.

Highlights

- Prognostic models for stage I and stage IV NSCLC were validated using data from 23 CALGB trials
- For stage I, Blanchon (B), Chansky (C) and Gail (G) models were evaluated
- For stage IV, Blanchon (B) and Mandrekar (M) models were evaluated
- Some models have better concordance with survival than the PS only model, but the absolute improvement is small

Figure 1.
Survival Curves of Risk Groups in All Patients

Table 1

Prognostic Models Investigated

	B model	C model	G model	M model
Citation	Blanchon et al (2006)	Chansky et al (2009)	Gail et al (1984)	Mandrekar et al (2006)
Geography Area	France	IASLC	LCSG	North America
Stage	stage I-IV	stage I-IIIa resected	resected stage I	advanced stage
Training n	2929	9137	392	782 from NCCTG
Validation n	1500	SEER 9221	na	433 from SWOG
Endpoint	OS	OS	Recurrence, OS	OS, TTP

OS: overall survival; TTP: time to progression; LCSG: Lung Cancer Study Group (LCSG); NCCTG: North Central Cancer Treatment Group; SWOG: Southwest Oncology Group

Table 2

Prognostic model risk score

Prognostic Model	Risk Score
B model	$1 \times (\text{if age} > 70) + 1 \times (\text{if gender} = \text{Male}) + 3 \times (\text{if PS} = 1) + 5 \times (\text{if PS} = 2) + 8 \times (\text{if PS} = 3) + 10 \times (\text{if PS} = 4) + 2 \times (\text{if histology} = \text{Large-cell}) + 3 \times (\text{if stage} = \text{IIA-IIIB}) + 6 \times (\text{if stage} = \text{IIIA-IIIB}) + 8 \times (\text{if stage} = \text{IV})$
C model	$\log(1.35) \times (\text{if Adenocarcinoma}) + \log(1.161) \times (\text{if Squamous}) + \log(1.38159) \times (\text{if Large cell}) + \log(1.353958) \times (\text{if Adenosquamous}) + \log(1.21) \times (\text{if gender} = \text{Male}) + \log(1.51) \times (\text{if age} > 70) + \log(1.30) \times (\text{if stage} = \text{IB}) + \log(1.872) \times (\text{if stage} = \text{IIA}) + \log(2.4336) \times (\text{if stage} = \text{IIB}) + \log(3.553056) \times (\text{if stage} = \text{IIIA})$
G model ^I	$-8.971 + 0.679 \times (\text{if T1N0 Nonsquamous}) + 0.103 \times (\text{if T1N1 Squamous}) + 1.600 \times (\text{if T1N1 Nonsquamous}) + 0.981 \times (\text{if T2N0 Squamous}) + 1.282 \times (\text{if T2N0 Nonsquamous}) + 0.450 \times (\text{if PS} = 2)$
M model	$0.26 + 0 \times (\text{if PS} = 0) + 0.48 \times (\text{if PS} = 1) + 0.96 \times (\text{if PS} = 2 \text{ or } 3) + 0.60 \times (\text{if underweight}) + 0 \times (\text{if normal weight}) + 0.11 \times (\text{if overweight}) - 0.11 \times (\text{if obese}) + 0 \times (\text{if normal HGB}) + 0.41 \times (\text{if abnormal HGB}) + 0 \times (\text{if normal WBC}) + 0.35 \times (\text{if high WBC})$

(if x) is an indicator function which equals 1 if the statement 'x' is true and 0 otherwise.

^IData on wound infection was not collected on 80% stage I NSCLC patients in the validation cohort. To evaluate the G model, we assumed all patients had no wound infection or equivalently the regression coefficient of infection is zero.

Table 3

Patient Characteristics for Model Validation

	Stage I (N=2060)	Stage IV (N=1611)
Age (yrs)		
n	2060	1611
Median	67.9	62.6
Age Group		
< 70	1194 (58.0%)	1259 (78.2%)
70	866 (42.0%)	352 (21.9%)
Sex		
Male	1102 (53.5%)	1022 (63.4%)
Female	958 (46.5%)	589 (36.6%)
Race		
White	1773 (86.1%)	1338 (83.1%)
Black	130 (6.3%)	191 (11.9%)
Other	156 (7.6%)	81 (5.0%)
Histology		
Adenocarcinoma	1001 (50.4%)	894 (55.8%)
Squamous	595 (30.0%)	297 (18.5%)
Large Cell	82(4.1%)	143 (8.9%)
Other	308 (15.5%)	268 (16.7%)
Performance Status		
0	1326 (64.7%)	565 (35.1%)
1	648 (31.6%)	867 (53.8%)
2	66 (3.2%)	177 (11.0%)
3/4	9 (0.4%)	2 (0.2%)
Weight Loss		
< 5%	1497 (88.8%)	650 (65.8%)
5%	189 (11.2%)	338 (34.2%)
WBC $\times 10^9/L$		
n	403	1238
Median	7.8	8.7
HGB g/dL		
n	557	1110
Median	13.4	13.0
BMI kg/m^2		
n	1020	1519
Median	26.4	25.0
Deaths	914 (44.4%)	1513 (93.9%)

Table 4

Survival Analysis Results for All Age Groups

Stage	Prognostic Model	n (Events)	Log Rank (p-value)	Correct Order?	HR (95%CI)
Stage 1	Blanchon	1977 (860)	128.90 (<.0001)	yes (4,3,2,1)	1.44 (1.35, 1.54)
	Chansky	1662 (755)	67.76 (<.0001)	yes (3,2,1)	1.13 (1.09, 1.16)
	Gail	1973 (856)	13.56 (0.0011)	no (3,1,2)	1.00 (0.97, 1.04)
	PS only	2049 (907)	107.30 (<.0001)	yes (2,1,0)	1.24 (1.19, 1.29)
Stage 4	Blanchon	1602 (1506)	105.47 (<.0001)	yes (6,5,4)	1.36 (1.30, 1.43)
	Mandrekar	1084 (1043)	159.86 (<.0001)	yes (5,4,3,2)	1.24 (1.20, 1.28)
	PS only	1611 (1513)	128.29 (<.0001)	yes (2,1,0)	1.20 (1.16, 1.24)

All risk scores standardized to the range [0, 10] for HR comparison.

Table 5

c-index and its Difference Relative to PS only Model for All Age Groups

Stage	Prognostic Model	c (95% CI)	c-difference (95% CI)	IDI (95% CI)
All Patients				
Stage 1	Blanchon	0.61 (0.59, 0.63)	0.034 (0.023, 0.045)	0.029 (−0.044, 0.061)
	Chansky	0.57 (0.55, 0.60)	−0.002 (−0.034, 0.026)	−0.027 (−0.080, 0.059)
	Gail	0.52 (0.50, 0.54)	−0.057 (−0.085, −0.029)	−0.053 (−0.561, 0.028)
	PS only	0.58 (0.56, 0.60)		
Stage 4	Blanchon	0.61 (0.60, 0.63)	0.015 (0.008, 0.023)	0.013 (0.007, 0.021)
	Mandrekar	0.64 (0.62, 0.66)	0.033 (0.020, 0.046)	0.012 (0.001, 0.020)
	M-B ^I		0.018 (0.003, 0.033)	−0.004 (−0.019, 0.008)
	PS only	0.60 (0.58, 0.61)		

^I Comparison between Mandrekar and Blanchon models.